

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model
Run on: January 31, 2002, 18:43:32 ; Search time 1929.4 Seconds
(without alignments)
3191.319 Million cell updates/sec

Title: US-09-641-831-1
Perfect score: 573
Sequence: 1 atgatgaggaccactgaaga.....gagcctattaagtatga 573

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 537289281 residues
Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database :
- 1: em_estfun:*
 - 2: em_esthum:*
 - 3: em_estin:*
 - 4: em_estom:*
 - 5: em_estpl:*
 - 6: em_estba:*
 - 7: em_estro:*
 - 8: em_estov:*
 - 9: em_htc:*
 - 10: gb_estli:*
 - 11: gb_est2:*
 - 12: gb_htc:*
 - 13: gb_gss:*
 - 14: em_gss_fun:*
 - 15: em_gss_hum:*
 - 16: em_gss_inv:*
 - 17: em_gss_pln:*
 - 18: em_gss_pro:*
 - 19: em_gss_rod:*
 - 20: em_gss_vrt:*
 - 21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	253	44.2	605	10 AW369663	AW369663 QV0-BT004
C 2	246.6	43.0	437	11 BF591083	BF591083 7h52f04.x
C 3	238	41.5	359	10 AW294613	AW294613 UI-H-B12-
C 4	204	35.6	300	10 AW590275	AW590275 hg42f03.x
C 5	158.4	27.6	265	10 AI300504	AI300504 qo19c11.x
C 6	150	26.2	649	11 BG563384	BG563384 602582446
7	100.8	17.6	1221	12 AK018481	AK018481 Mus muscu
8	91.2	15.9	321	10 BS504666	BS504666 BB504666
9	84.8	14.8	484	10 AU170966	AU170966 AU170966
10	84.6	14.8	481	11 BF463332	BF463332 UI-H-CS0P
11	82.6	14.4	231	10 AV368091	AV368091 AV368091
12	75	13.1	597	11 BF057621	BF057621 7k46h06.x

13	70.8	12.4	558	11 BF434691	BF434691 7p03b0..
14	66.6	11.6	354	10 AW001038	AW001038 wr91h03..
15	64.6	11.3	383	13 AQ244687	AQ244687 HS_2045.A
16	53.6	9.4	932	13 CNS038P4	AL258817 Tetraodon
17	52.4	9.1	614	13 AZ761049	AZ761049 1M0555J09
18	44.2	7.6	866	13 CNS05R29	AL350010 Tetraodon
19	43.6	7.6	1092	13 CNS0372N	AL231980 Tetraodon
20	40.6	7.1	379	10 AW356254	AW356254 38503 MAR
21	39.8	6.9	378	11 BF003124	BF003124 7g55602.x
22	39.4	6.8	586	13 AQ897936	AQ897936 HS_5536.B
23	39	6.8	412	11 W24504	W24504 zb61a09.r1
24	38.4	6.7	334	10 AA259846	AA259846 va36g01.r
25	38.4	6.7	553	10 AW845396	AW845396 CM4-CT004
26	38.4	6.7	1553	12 AK016447	AK016447 Mus muscu
27	38.2	6.7	458	10 AI825748	AI825748 wb69B08.x
28	38.2	6.7	882	11 BF693457	BF693457 602082111
29	37.8	6.6	387	10 AW058549	AW058549 wx23c10.x
30	37.4	6.5	290	10 AU168952	AU168952 AU168952
31	37.4	6.5	437	10 AI369738	AI369738 qv72406.x
32	37.4	6.5	465	10 AI309314	AI309314 qo80B08.x
33	37.2	6.5	398	10 AA111914	AA111914 zh53G03.F
34	37.2	6.5	473	10 AW083037	AW083037 xc05F08.x
35	37.2	6.5	924	10 BE730550	BE730550 601562213
36	37.2	6.5	946	13 CNS02DER	AL192348 Tetraodon
37	37	6.5	521	11 BF016403	BF016403 uv40G09.Y
38	36.8	6.4	185	10 AW230070	AW230070 up25B06.Y
39	36.8	6.4	288	11 W12275	W12275 ma06B09.r1
40	36.8	6.4	371	11 W20834	W20834 mb93a02.r1
41	36.8	6.4	479	10 AA855804	AA855804 vw75a10.F
42	36.8	6.4	499	10 AA764505	AA764505 vp05a05.F
43	36.8	6.4	515	10 AW743037	AW743037 up62a05.Y
44	36.8	6.4	532	10 AI509426	AI509426 va36g01.Y
45	36.8	6.4	558	10 AA869064	AA869064 vq30c11.F

ALIGNMENTS

RESULT 1
AW369663/c 605 bp mRNA EST 04-FEB-2000
LOCUS QV0-BT0041-011199-039-ell BT0041 Homo sapiens cDNA, mRNA sequence.
DEFINITION AW369663
ACCESSION AW369663
VERSION AW369663.1 GI:6874317
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 605)
AUTHORS HCGP <http://www.ludwig.org.br/ORESTES>.
TITLE The FAPESP/LICR Human Cancer Genome Project
JOURNAL Unpublished (1999)
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(<http://www.ludwig.org.br/scripts/gethtml2.pl?tti=QV0&st2=QV0-BT0041-011199-039-ell&t3=1999-11-01&t4=1>)
Seq primer: puc 18 forward
High quality sequence start: 17
High quality sequence stop: 586.
Location/Qualifiers
1. .605
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="BT0041"

FEATURES
source

```

/lab_host="DH10B"
/note=Organ: colon; Vector: pRT73D-Pac (Pharmacla) with a
modified polylinker; Site_1: Not 1; Site_2: Eco RI;
Plasmid DNA from the normalized library NCI_CGAP_C010 was
prepared, and ss circles were made in vitro. Following HAP
purification, this DNA was used as tracer in a subtractive
hybridization reaction. The driver was PCR-amplified cDNAs
from a pool of 5,000 clones made from the same library
(cloneids 1057416-1061255, and 1144584-1145351).
Subtraction by Bento Soares and M. Fatima Bonaldo.
117 a 100 c 101 g 119 t

```

Query Match	43.0%;	Score	246.6;	DB 11;	Length	437;			
Best Local Similarity	98.4%;	Pred.	No. 2.3e-60;						
Matches	249;	Conservative	0;	Mismatches	4;	Indels	0;	Gaps	0;
<hr/>									
QY	321	cagcctctcccccacgcggtcacctgttagcctcgagttagacctcaacatccccacaggat	380						
Dbb									
Dbb	366	CAGCCTCTCCCCACCGGTCACTGTTAGCCTCGAGTTTGACCCTCACATCCCCACAGGAT	307						
Dbb									
QY	381	gcctcctaagaactcgaaacctattttagttctctetactgtctgatggacgcatagtggs	440						
Dbb									
Dbb	306	GCCTCCTAGGACTCGAACCTCATTTTAGTGTCTCTACTGTGTATGGACGCCATGAGTGGAG	247						
Dbb									
QY	441	ctgtcgaccacaccttgggtgaagtgggtgtccacgtgccacctgggcagcagcatggcc	500						
Dbb									
Dbb	246	CTGTGACACCACTTGGGTGAAGTGTGTCTCCACGTCCACCCTGGCAGACGATGGCC	187						
Dbb									
QY	501	acagaaaaggttgttatctataccccaccacccagcacacaacatgcagaaaaattccaagaagccta	560						
Dbb									
Dbb	186	ACAGAAAGGTTTATTATTACCCCAGCGCCCAACACATGCAGAAAATTTCAA AAAAGCCTTA	127						
Dbb									
QY	561	tttaagtagatatga	573						
Dbb									
Dbb	126	TTTAAGTAGATCA	114						

RESULT 3
AW294613/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

359 bp mRNA
UI-H-B12-ahk-h-11-0-UI-s1 NCI_CGAP_Sub4 Homo sapiens cDNA clone
IMAGE:2727404 3', mRNA sequence.
AW294613
AW294613
AW294613.1 GI:6701249
EST.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 359)
NCI-CCGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.

Contact: crapbs-r@mail.nih.gov
 Email: robert.strausberg.phd.D
 The sequence contained an oligo-qt track that was present in the
 oligonucleotide that was used to prime the synthesis of first
 strand cDNA and therefore this may represent a bonafide poly A
 tail. cDNA Library Preparation: M.B. Soares Lab Clone distribution:
 NCI-CGAP clone distribution Information can be found through the
 I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
 Seq primer: M13 Forward
 POLYA=Yes.

```

FEATURES
  source
    Location/Qualifiers
      1..359
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /clone IMAGE:2727404"
        /clone_lib="NCI CGAP Sub4"

```


REFERENCE 1 (bases 1 to 265)
 AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 JOURNAL Tumor Gene Index
 COMMENT Unpublished (1997)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
 Emmert-Buck, M.D., Ph.D.

CDNA Library Preparation: M. Bento Soares, Ph.D.
 CDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.lnl.gov/db/ncicgap/ncicgap.html
 Insert Length: 639 Std Error: 0.00
 Seq primer: -40UP from Glbco
 High quality sequence stop: 219.
 Location/Qualifiers

FEATURES
 source
 1..265

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:1908980"
 /clone_lib="NCI CGAP Lys"
 /tissue_type="carcinoid"
 /lab_host="DH10B"
 /note="Organ: Lung; Vector: pT73D-Pac (Pharmacia) with a
 modified polylinker; 1st strand cDNA was prepared from a
 neuroendocrine lung carcinoid, and was then primed with a
 Not I - oligo(dT) primer. Double-stranded cDNA was ligated
 to Eco RI adaptors (Pharmacia), digested with Not I and
 cloned into the Not I and Eco RI sites of the modified
 pT73 vector. Library is normalized. Library was
 constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 75 a 61 c 50 g 79 t
 ORIGIN

Query Match 27.6%; Score 158.4; DB 10; Length 265;
 Best Local Similarity 99.4%; Pred. No. 4.6e-35;
 Matches 159; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 414 taccgtcatgagccatgagtgagctgtcgaccaccccttgggtgaagtgtgtctcc 473
 |||||||
 DB 265 TACTGCTATGAGCCGCTAGTGAGTGTGCGACACCTTGGGGAAGTGTGTCTCC 206

QY 474 acgtccacctgtggcagcagatggccacagaagaagtgtgtcttaccacccacacaa 533
 |||||||
 DB 205 ACGTCCACCTGGGCGAGCAGCATGGCCACAGAAAGTGTATTACCCACCCAGACAA 146

QY 534 cacatgcagaaattcaaaagacgtatttaagtatga 573
 |||||||
 DB 145 CACATGCAGAAATTTCANAAAGAGCTATTTAAGTAGATGA 106

RESULT 6
 LOCUS BG563384 649 bp mRNA EST 10-APR-2001
 DEFINITION 6035824461 NIH_MGC_76 Homo sapiens CDNA clone IMAGE:4710299 5',
 mRNA sequence.

ACCESSION BG563384
 VERSION BG563384.1 GI:13571036
 KEYWORDS EST.
 SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 649)
 AUTHORS NIH-MGC <http://mhc.nci.nih.gov/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov

Tissue Procurement: CLONTECH Laboratories, Inc.
 CDNA Library Preparation: CLONTECH Laboratories, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LHC1548 row: 1 column: 12
 High quality sequence stop: 548.
 Location/Qualifiers

FEATURES
 source

1..649
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4710299"
 /clone_lib="NIH MGC 76"
 /lab_host="DH10B (TI phage-resistant)"
 /note="Organ: Liver; Vector: pDNR-LIB (Clontech); Site: 1;
 Site 1 (ggcgcctcgcgc); Site 2: Site 1 (ggcgcctcgcgc); 5' and
 3' adaptors were used in cloning as follows: 5' adaptor
 sequence: 5'-CACGCCATTATGGC-3' and 3' adaptor sequence:
 5'-ATTCTAGAGCGCGAGCGCGCGACATG-(30)BN-3' (where B = A,
 C, or G and N = A, C, G, or T). Average insert size 1.85
 kb (range 1.0-4.0 kb). 15/15 colonies contained inserts
 by PCR. This library was enriched for full-length clones
 and was constructed by Clontech Laboratories (Palo Alto,
 CA). Note: this is a NIH-MGC Library."

BASE COUNT 161 a 172 c 193 g 123 t
 ORIGIN

Query Match 26.2%; Score 150; DB 11; Length 649;
 Best Local Similarity 96.5%; Pred. No. 1.6e-32;

Matches 164; Conservative 0; Mismatches 5; Indels 1; Gaps 1;

QY 163 tctaaagtcgaagaaggcgaagaacagacaccctgaagctccaactcaggtcggat 222
 |||||||
 DB 90 TATAGCTGAGAGAGGGGCGCAAGCAGACACCTCGAGCTCCAACTGAGCTGGGAT 149

QY 223 gaagttgtcacatcaatgaagtgactctgagcagctccagaagaaggcagttccctg 282
 |||||||
 DB 150 GAGGTGTGACATCAATGAGTGACTCTGACACGCTCCAGAAAGAGGACGTTCCCTG 209

QY 283 gtaaaagatccctacaagaaccctcagg-ctgttagtgcgcacctctccc 331
 |||||||
 DB 210 GTGAAGGATCTTACAAGACCTCTAGGACTGTAGTCCGACAGGTCCTCCG 259

RESULT 7

LOCUS AK018481 1121 bp mRNA HTC 05-JUL-2001
 DEFINITION Mus musculus 16 days embryo lung CDNA, RIKEN full-length enriched
 library, clone:8430440M04, full insert sequence.

ACCESSION AK018481
 VERSION AK018481.1 GI:12858200
 KEYWORDS CAP trapper.
 SOURCE Mus musculus (strain:C57BL/6J) 16 days embryo lung CDNA to mRNA,
 clone_lib:RIKEN full-length enriched mouse CDNA library
 clone:8430440M04.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 1121)
 AUTHORS Carninci, P. and Hayashizaki, Y.

TITLE High-efficiency full-length cDNA cloning

JOURNAL Methods in enzymology. 303, 19-44 (1999)

PUBMED 99279253

REFERENCE 2 (bases 1 to 1121)

AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
 Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
 TITLE Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new genes
 JOURNAL Genome research. 10 (10), 1617-1630 (2000)


```

clone_1lb-"RIKEN full-length enriched, 0 day neonate
kidney"
/tissue_type="kidney"
/dev_stage="0 day neonate"
/lab_host="DH10B"
/note="Site_1: Sall; Site_2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN, Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5',
GAGGAGAGAGAGCGGCCGCACACTCGAGATTTTTTTTTTTTNN 3'}, cDNA was
prepared by using reverse transcriptase thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. Second strand cDNA was prepared with the
primer adaptor of sequence [5',
GAGGAGAGAGATTTCGAGATTAAATTAATATCCCCCCCCCCCCC 3'}. cDNA
was cleaved with BamHI and XhoI. Vector: a modified
phluescript KS(+) after bulk excision from Lambda PLoC I."

```

Query Match	15.98;	Score 91.2;	DB 10;	Length 321;
Best Local Similarity	71.48;	Pred. No. 1.1e-15;		
Matches 120;	Conservative 0;	Mismatches 48;	Indels 0;	Gaps 0;

0y 3 gatggagggccacattaaagcttcacaaagccatgagcgcaattaaactcaaacacggcac 62
db 76 gatgaagattccggagaaacttgagaggaacctgactccaccccccaccccatgtaggacccc 135
0y 63 caaaggaaaggtacattctctcggagagccttcctcggagaggaagacctccctgggtttac 132
db 136 caccgagagattccgttcattctgaaagccctctctgagaggaaggagcgtcccccgggcttcac 195

```

Oy      123 tctaaggtgctgcctgagcagcagaacccatlaatcatctctaagt 170
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      196 CTGAAGGTGGCCTGGAGCGTGGAGAACCGTATATCATTTCCAGGT 243

```

RESULT 9
AU170966

LOCUS	484 bp	mRNA	EST	29-JAN-2001
DEFINITION	AU170966	ol-br-ad cDNA	Oryzias latipes	cDNA clone br7042, mRNA sequence.

ACCESSION	AUI/0966
VERSION	AUI70966.1
KEYWORDS	GI:12593037
SOURCE	EST.
	Japanese medaka.

ORGANISM *Oryzias latipes*
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;

REFERENCE
Mita, K., Ishikawa, Y. and Yamauchi, M.
1 (bases 1 to 484)

TITLE	Establishment of cDNA database of medaka, <i>Oryzias latipes</i>
JOURNAL	Unpublished (2001)
CONTACT	Contact: Mita K
COMMENT	Genome Research Group

Genome Research Group
National Institute of Radiological Sciences
Anagawa 4-9-1, Inage, Chiba 263-8555, Japan
Email: kmkita@nirs.ri.go.jp

```
method: uni-directional sequence direction: sequenced from T3 primer
(5' -> 3').
location/Qualifiers
    source          1
    length          104
```

SOURCE
1. 484
/organism="Oryzias latipes"
/strain="HNI"
/db_xref="taxon:8090"
/clone="br042"
/clone_1lb="O1-br-ad cDNA"

```

/sex="female/male mixed"
/tissue_type="brain"
/dev_stage="adult"
BASE COUNT      130 a      150 g      90 t
ORIGIN           114 c

```

Query Match	14.88;	Score 84.8;	DB 10;	Length 484;
Best Local Similarity	65.98;	Pred. No. 8.5e-14;		
Matches 141; Conservative	0;	Mismatches 67;	Indels 6;	Gaps 1.

Qy 100 ggaagagctccctggttttaactctaagggtgcctggagcacgagaacatcaatc 159
|| || || || || || || || || || || || || || || || || || || || || ||
Db 277 GGGGCGCGGCATGGGGTTTACCCCTCAGAGAGAGACTGTGAGCATCAGAACCGCTTGTC 336

160 atcctaaggtcgaaaggggcaaacagacacccctagctccaactgcagctggg 219

DB 33/ ATATCCAAAGTGGAGGAGGCGACGAAACCGGCACCGTAGA-----CTTCAGTGGGA 279

QY 220 gatgaggttctgcacatcaatgatgagtgaactctgagcagctccagaaggagcgattcc 279

Db 391 GACGAGCTCTCACATTAACCCAGATTCCCTGAGGGGCTACCGGCAGAGGCCATCTGC 450

QY 280 ctggtgaagatcctacaagaccctcaagcttg 313

Db 451 CTGGTGAAGGCGCTCCACACAGACCCCTGACTCTGG 484

	RESULT	10		
	BFA63332			
LOCUS	BFA63332	481 bp	mRNA	EST 04-DEC-2000
DEFINITION	NF-W-CG96-phm-mc205-D[UTR]		s1 NIH BMAD Pat4 S2 Mus musculus cDNA clone	

DATE INITIATION 01-M-0609p
 UI-M-CG09p.bnm-c-05-0-01 3, mRNA sequence.
 ACCESSION BF463332
 VERSION BF463332.1 GI:11532515

KEYWORDS	EST.
SOURCE	house mouse.
ORGANISM	Mus musculus
	Eukaryota: M

REFERENCE
AUTHORS
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus
1 (bases 1 to 481)
Bonaldi, M.F., Lennon, G. and Soares, M.B.

TITLE Normalization and subtraction: two
discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477

COMMENT Contact: Chin, H
National Institute of Mental Health
6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD

20892-9643, USA
Tel: 301 443 1706
Fax: 301 443 9890
Email: WEST@mail.sri.com

Oligo-dT track not found. Not I site shown in beginning of sequence is likely internal to the message. cDNA library preparation: M.B. Soares Lab Clone distribution: Researchers may obtain BMAP cDNA

clones from RESEARCH GENETICS. It should be noted that Benclo Soares is generating a small number of additional specialized non-redundant arrays of BMAP CDNA's whose availability will be considered under appropriate and limited collaborative arrangements.

Seq primer:	M13 Forward	Location/Qualifiers
POLYA-No.		

FEATURES

```

source      1. .481
            /organism="Mus musculus"
            /strain="C57BL/6J"
            /db xref="taxon:10090"

```

```

/clone="UI-M-CG0p-dam-c-05-0-0-UI"
/clone_lib="NIH_BMAP_Ret4_S2"
/lab_host="DH10B (Life Technologies)"

```

```
/note=Vector: pT7T3D-Pac (Pharmacia) with a mod  
polylinker, Site_1: Not I; Site_2: Eco RI; The  
NIH_BMAP_Ret4_S2 library is a subtracted library
```


cDNA Library Preparation: M. Bento Soares, Ph.D.
 cDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
 www-bio.llnl.gov/bbrp/image/Image.html
 Insert Length: 330 Std Error: 0.00
 Seq primer: -40UP from Gibco.

FEATURES

Location/Qualifiers
 source

1..354
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:2495093"
 /clone_lib="NCI_CGAP_Kid11"
 /lab_host="DH10B"
 /note="organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with
 a modified polylinker; Site 1: Not I; Site 2: Eco RI;
 Plasmid DNA from the normalized library NCI_CGAP_Kid3 was
 prepared, and ss circles were made in vitro. Following HAP
 purification, this DNA was used as tracer in a subtractive
 hybridization reaction. The driver was PCR-amplified cDNAs
 from a pool of 5,000 clones made from the same library
 (cloneIDs 1322376-1323911, 1456007-1456775, and
 1500552-1502855). Subtraction by Bento Soares and M.
 Fatima Bonaldi."

BASE COUNT 82 a 85 c 80 g 107 t

ORIGIN

Query Match 11.6%; Score 66.6; DB 10; Length 354;
 Best Local Similarity 60.8%; Pred. No. 1.3e-08;
 Matches 144; Conservative 0; Mismatches 89; Indels 4; Gaps 2;
 QY 111 ctgggggtttactctaaaggtggtgagcagcgagagacattatca-tctctaaagg 169
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 354 CTGGGGCTTCACCTTAAGGGGGTCTGGAACCTCTGAAACCGCTCCCGCTCTAAGA 295
 QY 170 tcgaagaaggggcaagacagacacctgagctccaaactgcaggctgggagtggtg 229
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 294 TTGAAGATGGAGGCAAGGCAG---CTTTGTCCAGAGATGAGGACTGGTGAAGCTGG 238
 QY 230 tgcacatcaatgaggtgactctgagcagctccagaaaggaggttccctggtgaaag 289
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 237 TGAATATCATGGCACCTCCATATATGCTCCGCCAAGAGGCCCTCATCTCATCAAG 178
 QY 290 gatctacaagacctcaggtgtagtgcgagcctctcccccacccggtcactgtta 346
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 177 GCTCCTCCGGATTCTCAAGCTGATTGTCAGGAGACTAAGTTTGTGCGCCCTGGTA 121

RESULT 15

AQ244687 383 bp DNA GSS 03-OCT-1998
 LOCUS HS_2045_A2_H11.T7 CIT Approved Human Genomic Sperm Library D Homo
 DEFINITION sapiens genomic clone Plate=2045 Col=22 Row=O, DNA sequence.
 ACCESSION AQ244687
 VERSION AQ244687.1 GI:3691261
 KEYWORDS GSS.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 383)
 Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
 Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
 Hood,L.

TITLE Sequence-tagged connectors: A sequence approach to mapping and
 scanning the human genome
 JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
 MEDLINE 99380589
 COMMENT Contact: Mahairas GG, Wallace JC, Hood L
 High Throughput Sequencing Center
 University of Washington

401 Queen Anne Avenue North, Seattle, WA 98109, USA
 Tel: (206) 616-3618
 Fax: (206) 616-3887
 Email: jwallace@u.washington.edu
 Sequence Tagged Connector
 Plate: 2045 row: O column: 22
 Class: BAC ends
 High quality sequence stop: 383.

FEATURES

Location/Qualifiers
 source

1..383
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="plate=2045 Col=22 Row=O"
 /clone_lib="CIT Approved Human Genomic Sperm Library D"
 /sex="male"
 /note="Organ: sperm; Vector: pBelobAC11; BAC Clones in
 E-Coli DH10B; 115 g 90 t
 BASE COUNT 86 a 92 c 115 g 90 t

ORIGIN

Query Match 11.3%; Score 64.6; DB 13; Length 383;
 Best Local Similarity 65.7%; Pred. No. 5.1e-08;
 Matches 94; Conservative 0; Mismatches 49; Indels 0; Gaps 0;
 QY 175 gaagggggcaagcagacacccctgagctccaaactgcaggctgggagtgaggttgcac 234
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 216 GAAGAGGGCAGTAAAGCGCGGCTCGACAAAGTTACTGTGGAGATGAGATCGTCGGC 275
 QY 235 atcaatgaggtgactctgagcagctccagaaaggaggttccctggtgaaaggtacc 294
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 276 ATCAATGACATTGGTCTCTCAGGGTTTAGACAGGAAGCGATTTCCTGCTGAAGGGTCC 335
 QY 295 tacaagaccctcaggctggtagt 317
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 336 CATAGACCCCTGAAGCTGTCGT 358

Search completed: January 31, 2002, 19:16:09
 Job time: 1957 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 31, 2002, 19:16:09 ; Search time 1929.4 Seconds
(without alignments)
1821.224 Million cell updates/sec

Title: US-09-641-831-3
Perfect score: 327
Sequence: 1 atgatgagaccactgaaga.....ggctgtagtgcgagttga 327

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: em_estfun:*
2: em_esthum:*
3: em_estin:*
4: em_estom:*
5: em_estpl:*
6: em_estba:*
7: em_estro:*
8: em_estov:*
9: em_htc:*
10: gb_estl1:*
11: gb_estl2:*
12: gb_htc:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pin:*
18: em_gss_pro:*
19: em_gss_rod:*
20: em_gss_vrt:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	148.4	45.4	649	11	BG563384
2	100.8	30.8	1121	12	AK018481
3	91.2	27.9	321	10	BB504666
4	84.8	25.9	484	10	AU170966
5	84.6	25.9	481	11	BF463332
6	82.6	25.3	231	10	AV368091
7	77.5	22.9	597	11	BF057821
8	70.8	21.7	558	11	BF434691
9	65	19.9	383	13	AQ244687
10	62.8	19.2	354	10	AW001038
11	54	16.5	932	13	CNS03SP4
12	45	13.8	866	13	CNS05R29

13	44.4	13.6	1092	13	CNS0372N
14	42.2	12.9	458	10	AI825748
15	40.6	12.4	379	10	AW356254
16	39.4	12.0	586	13	AQ697936
17	39	11.9	412	11	W24504
18	38.4	11.7	334	10	AA259846
19	38.4	11.7	553	10	AW845396
20	38.4	11.7	1553	12	AK016447
21	38.2	11.7	882	11	BF693457
22	37.4	11.4	290	10	AU168952
23	37.2	11.4	398	10	AA111914
24	37	11.3	378	11	BF003124
25	37	11.3	521	11	BF016403
26	36.8	11.3	185	10	AW230070
27	36.8	11.3	288	11	W12275
28	36.8	11.3	371	11	W20834
29	36.8	11.3	479	10	AA855804
30	36.8	11.3	499	10	AA764505
31	36.8	11.3	515	10	AW743037
32	36.8	11.3	532	10	AI509426
33	36.8	11.3	558	10	AA869064
34	36.8	11.3	581	11	BG922902
35	36.6	11.2	212	10	AL603379
36	36.6	11.2	300	10	AU098787
37	36.6	11.2	309	11	D31139
38	36.6	11.2	321	10	AA374357
39	36.6	11.2	323	11	BG407069
40	36.6	11.2	375	10	AA112776
41	36.6	11.2	415	10	AA329300
42	36.6	11.2	455	10	AA446635
43	36.6	11.2	456	11	H72671
44	36.6	11.2	463	11	BF001694
45	36.6	11.2	467	11	W07582

ALIGNMENTS

RESULT 1
BG563384 649 bp mRNA 10-APR-2001
LOCUS 602582446F1 NIH_MGC_76 Homo sapiens cDNA clone IMAGE:4710299 5',
DEFINITION mRNA sequence.
ACCESSION BG563384
VERSION BG563384.1 GI:13571036
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 649)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs@email.nih.gov
Tissue Procurement: CLONETECH Laboratories, Inc.
CDNA Library Preparation: CLONETECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM1548 row: 1 column: 12
High quality sequence stop: 548.
Location/Qualifiers
1. .649
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4710299"
/clone_lib="NIH_MGC_76"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: liver; Vector: pDNR-LIB (Clontech); Site: 1."

Db 337 ATACCAAGGTGAGGAGGACCAAGCGCGCTGAGT-----CTCAAGTGGGA 390
 Oy 220 gatgaggtgtgtacatcaatgagtgactcttagagcgtccagaagaggaagcattcc 279
 Db 391 GAGGAGCTCTCAACATTAACCCAGATTCCTCCCTGAGGGCTACCGGAGAGGCCATCTGC 450
 Oy 280 ctgtgtaaagagactcaagaagacctcaggctgg 313
 Db 451 CTGTGTAAGGGCTCCCAAGACCTGACTCTGC 484

RESULT 5
 BF463332 481 bp mRNA EST 04-DEC-2000
 LOCUS UI-M-CGDP-dnm-c-05-0-UI.s1 NIH_BMAP_Ret4_S2 Mus musculus cDNA clone
 DEFINITION UI-M-CGDP-dnm-c-05-0-UI 3', mRNA sequence.
 ACCESSION BF463332
 VERSION BF463332.1 GI:11532515
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 481)
 AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
 TITLE Normalization and subtraction: two approaches to facilitate gene discovery
 JOURNAL Genome Res. 6 (9), 791-806 (1996)
 MEDLINE 97044477
 COMMENT Contact: Chin, H
 National Institute of Mental Health
 6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
 20892-9643, USA
 Tel: 301 443 1706
 Fax: 301 443 9890
 Email: mestr@mail.nih.gov
 Oligo-dT track not found. Not 1 site shown in beginning of sequence
 is likely internal to the message. cDNA library preparation: M.B.
 Soares Lab Clone distribution: Researchers may obtain BMAP cDNA
 clones from RESEARCH GENETICS. It should be noted that Bento Soares
 is generating a small number of additional specialized
 non-redundant arrays of BMAP cDNAs whose availability will be
 considered under appropriate and limited collaborative arrangements
 Seq primer: M13 forward
 POLYA-No.

FEATURES
 source Location/Qualifiers

1. 481
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UI-M-CGDP-dnm-c-05-0-UI"
 /clone_1lb="NIH_BMAP_Ret4_S2"
 /lab_host="DH10B (Life Technologies)"
 /note="Vector: pRT3D-Pac (Pharmacia) with a modified
 polylinker; Site_1: Not 1; Site_2: Eco RI; The
 NIH_BMAP_Ret4_S2 library is a subtracted library,
 ultimately derived from mouse retina tissue libraries at
 various stages of development. For a detailed description
 of the library from which this clone was derived, please
 visit our web site at brainest.eng.uiowa.edu.
 TAG_SEQ=None found"
 BASE COUNT 125 a 140 c 115 g 101 t
 ORIGIN

Query Match 25.9%; Score 84.6; DB 11; Length 481;
 Best Local Similarity 69.5%; Pred. No. 2.9e-13;
 Matches 130; Conservative 0; Mismatches 54; Indels 3; Gaps 1;
 Oy 131 gtggcctggaagcaagcattcaatcctcgaagtgcaaggaaggggcaagag 190
 Db 1 GCGGCGCGAGCATGCGACGCTCTGTCATCCACCAAGATTGAGGAAGCAAGAGCTG 60

Oy 191 acaccctagctccaaactcgaagctgggagtgaggtgtgtcacatcaatgagtgactc 250
 Db 61 CAGCAGTGGAC---AAGTACTTCTGAGATGAGATAGTGGCCATTATGATGTGATC 117
 Oy 251 tgaagagctccagaagagagcagttccctgggtgaagagatcctacaagacctcaggc 310
 Db 118 TCTCAGGGTTTCAGACAAAGACATTTTCCTGCTGTAAGAGCTCCCAAGACCTGAAGC 177
 Oy 311 tggtagt 317
 Db 178 TTGTGCT 184

RESULT 6
 AV368091 231 bp mRNA EST 14-NOV-1999
 LOCUS AV368091 RIKEN full-length enriched, 16 days embryo lung Mus
 DEFINITION Musculus cDNA clone 8430440M04 3', mRNA sequence.
 ACCESSION AV368091
 VERSION AV368091.1 GI:6415738
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 231)
 AUTHORS Komuro,H., Aizawa,K., Akahira,S., Akiyama,J., Carninci,P., Endo,T.,
 Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N., Hirozane,T., Hori,F.,
 Ishii,Y., Ishikawa,T., Itoh,M., Izawa,M., Kadota,K., Kagawa,I., Kai
 C., Kawai,J., Kikuchi,N., Kojima,Y., Koya,S., Kusakabe,M.,
 Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.,
 Owa,C., Ozawa,Y., Saito,H., Sano,M., Sato,K., Shibata,K., Shibata
 Y., Shigemoto,Y., Shiraki,T., Sogabe,Y., Sugihara,Y., Suzuki,H.,
 Suzuki,H., Takahashi,F., Tateo,M., Tomimaga,N., Tsunoda,Y.,
 Watahiki,A., Watanabe,S., Yamamuta,T., Yasunishi,A., Yokota,T.,
 Yoshiki,A., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
 RIKEN Mouse ESTs (Komuro,H., et al. 1999)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Yoshihide Hayashizaki
 Laboratory for Genome Exploration Research Group, RIKEN Genomic
 Sciences Center(GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Suenho-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9222
 Fax: 81-45-503-9216
 Email: genome-resgsc.riken.go.jp.
 URL:http://genome.gsc.riken.go.jp/
 Sasakhi,N., Izawa,M., Watanaki,M., Okazaki,Y. and Hayashizaki
 Y.
 Matsuyama,S., Carninci,P., Muramatsu,M., Okazaki,Y. and Hayashizaki
 Y.
 Transcriptional sequencing: A method for DNA sequencing using RNA
 polymerase. Proc. Natl. Acad. Sci. U.S.A. 95 (7), 3455-3460 (1998)
 Itoh,M., Katsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
 Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki
 Y., and Hayashizaki,Y.
 Automated filtration-based high-throughput plasmid preparation
 system. Genome Res. 9 (3), 463-470 (1999)
 Carninci,P. and Hayashizaki,Y.
 High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
 19-44 (1999)
 Please visit our web site (http://genome.rtc.riken.go.jp) for
 further details.

FEATURES
 source Location/Qualifiers

1. 231
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="8430440M04"
 /clone_1lb="RIKEN full-length enriched, 16 days embryo
 lung"
 /sex="mixed"
 /tissue.type="lung"
 /dev_stage="16 days embryo"

/lab_host="DH10B"
/note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5',
GAGAGAGAGAGGATCCAAAGAGCTCTTTTCTTTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot - 10.0 and subtraction to Rot - 185.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATCTCGAGTAAATATCCCTCCCCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI"

BASE COUNT 52 a 56 c 52 g 71 t
ORIGIN

Query Match 25.3%; Score 82.6; DB 10; Length 231;
Best Local Similarity 71.2%; Pred. No. 8.8e-13;
Matches 109; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

QY 18 agacttcacaaagcctagtgccattataacttaacacagggccacaaagggaggtacat 77
|||||
Db 1 AAAATTGTAGCAGCTAGCCCTACTCAAAACCCCTACGACGCCACCCAGAGATTCCT 60
|||||
QY 78 ttatctgaggaattctgagagagagctccctgggttttactctaaagggtgacct 137
|||||
Db 61 TTATGTGAACCTCTCTAGAGGAGGGATCCCTGGGCTTACCTCAAGGTGGCCT 120
|||||
QY 138 ggagcagggagacattatcatctctaaagt 170
|||||
Db 121 GGAGCGTGAGACCGTATTATTCATTCCAAGGT 153
|||||

RESULT 7
BF057621 597 bp mRNA EST 16-OCT-2000
LOCUS
DEFINITION
7k4606.x1 NCI-CGAP_Ov18 Homo sapiens cDNA clone IMAGE:3478523 3', similar to SW:APXL_HUMAN Q13796 APICAL-LIKE PROTEIN ; contains MER22 b2 TARI repetitive element ;, mRNA sequence.

ACCSSION
BF057621
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 597)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: M. Bento Soares, Ph.D. cDNA Library Arrayed by: Christa Prange, The I.M.A.G.E. Consortium DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL, send email to: info@image.llnl.gov
Seq primer: -400P from Gibco
High quality sequence stop: 511.
Location/Qualifiers
1. .597
/organism="Homo sapiens"

FEATURES
source

/db_xref="taxon:9606"
/clone="IMAGE:3478523"
/clone_lib="NCI-CGAP_Ov18"
/tissue_type="fibrothecoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: ovary; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5',
TGTACCAATCTGAAGTGGAGCGCGCGGACATTTTTTTTTTTT 3'];
double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 125 a 205 c 158 t 108 t
ORIGIN

Query Match 22.9%; Score 75; DB 11; Length 597;
Best Local Similarity 66.3%; Pred. No. 1.4e-10;
Matches 124; Conservative 0; Mismatches 60; Indels 3; Gaps 1;

QY 131 gtggcctgagcagcagacattatcatcttaagtcgaagggggcgaagcag 190
|||||
Db 1 GCGGCGCGAGCAGCGCGCGCTGCTCATCCACAGATTGAAGAGGCGCAGTAAAGC-- 58
|||||
QY 191 acacctgagctccaaactgcagctgaggtgaggtgtgtgcacatcaatgagtgactc 250
|||||
Db 59 -CGCGCGCTCGACAAGTTACTTGGCTGGAGATGAGATCGCGCATCAATGACATTGGTC 117
|||||
QY 251 tgagcagctccagaagagagcagttccctgggtgaagagatcctacagaccctcaggc 310
|||||
Db 118 TCTCAGGTTTAGACAGGAGCGGATTCCTGGTGAGGGTCCCTAAGACCCCTGAGC 177
|||||
QY 311 tggtagt 317
|||||
Db 178 TGGTGTG 184
|||||

RESULT 8
BF434691 558 bp mRNA EST 29-NOV-2000
LOCUS
DEFINITION
7p03b06.x1 NCI-CGAP_Ov18 Homo sapiens cDNA clone IMAGE:3644507 3', similar to SW:APXL_HUMAN Q13796 APICAL-LIKE PROTEIN ;, mRNA sequence.

ACCSSION
BF434691
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 558)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: M. Bento Soares, Ph.D. cDNA Library Arrayed by: Christa Prange, The I.M.A.G.E. Consortium DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL, send email to: info@image.llnl.gov
Seq primer: -400P from Gibco
High quality sequence stop: 488.
Location/Qualifiers
1. .558
/organism="Homo sapiens"
/db_xref="taxon:9606"

FEATURES
source

BASE COUNT
ORIGIN

```
Query Match      19.2%; Score 62.8; DB 10; Length 354;
Best Local Similarity 62.1%; Pred. No. 3e-07;
Matches 133; Conservative 0; Mismatches 77; Indels 4; Gaps 2;

QY 111 ctggggttttactctaaaggtggtcgtgagcagcagagaccattatca-tctctaaag 169
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 354 CTGGGGCTTCAACCTTAAGGGGGTGTGAACCCCTGTGAACCGGCTCCAGTGTCTAAGA 295
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 170 tcgaagaaggggcaagcagacacctgagctccaaactgcaggtcggggtgaggttg 229
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 294 TTGAAGATGGAGCAGGAGCAG---CTTGTCCAGAGATGAGGACTGGTGTGATGAGCTGG 238
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 230 tgacatcaatgaggtgactctgagcagctccagaagggcaggttccctcgttgtaag 289
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 237 TGAATCATGATGCATCCATATATATGCTCCGCCAGAGAGGCCCTCATCTCATCAAG 178
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 290 gatctacaagaccctcaggtcgtggtgagtcgag 323
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 177 GTCCTTCGGATTCTCAAGCTGATTGTGAGGAG 144
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 11
CNS03SP4      932 bp      DNA      17-MAY-2000
LOCUS      Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone
DEFINITION      054119 of library G from Tetraodon nigroviridis, genomic survey
                sequence.
ACCESSION      AL258817.1 GI:7979829
VERSION      AL258817.1 GI:7979829
KEYWORDS      GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis.
ORGANISM      Tetraodon nigroviridis
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
                Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
                Tetraodontidae; Tetraodon.
REFERENCE      1 (bases 1 to 932)
AUTHORS      Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
                Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
                Weissenbach,J.
TITLE      Characterization and repeat analysis of the compact genome of the
                freshwater pufferfish Tetraodon nigroviridis
JOURNAL      Unpublished
REFERENCE      2 (bases 1 to 932)
AUTHORS      Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
                Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
                Saurin,W. and Weissenbach,J.
TITLE      Human gene number estimate provided by genome wide analysis using
                Tetraodon nigroviridis DNA sequence
JOURNAL      Unpublished
REFERENCE      3 (bases 1 to 932)
AUTHORS      Genoscope.
TITLE      Direct Submission
JOURNAL      Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
REFERENCE      This sequence is a single read and was generated as part of a large
                scale clone-end sequencing project of the Tetraodon nigroviridis
                genome. For more information, please take a look at
                http://www.genoscope.cns.fr/Tetraodon.
FEATURES      Location/Qualifiers
                source      1..932
                        /organism="Tetraodon nigroviridis"
                        /db_xref="taxon:99883"
                        /clone_lib="G"
                        /note="Genoscope sequence ID : COBG054AE10SP1-end :
                        PUC-Ori"
BASE COUNT      177 a 260 c 303 g 189 t 3 others
ORIGIN

Query Match      16.5%; Score 54; DB 13; Length 932;
Best Local Similarity 66.1%; Pred. No. 0.0001;
Matches 78; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

QY 194 ccctgagctccaaactcaggtcgggtgaggttggtgcacatcaatgaggtgactctga 253
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 227 CCGTGTGGCCAACTGAGGTGGCGCAGGCTCATCAATCAACGATCCACCTGT 168
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 254 gcagctccagaagaggcaggttccctcgttgaaagatccacagacccctcaggtcg 313
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 167 ACGCAGCAGCAGGAGAGCCCTNATCCTCATCAAGGGCTGTTCCGCATCCTGAAGTCA 108
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 314 tagtgcagcttga 327
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 107 CGGTGAGGAGGTAA 94
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

RESULT 13
 CENS0372N
 LOCUS
 DEFINITION CENS0372N 1092 bp DNA GSS 15-MAY-2000
 Tetraodon nigroviridis genome survey sequence PUC-01 end of clone
 004C04 of library G from Tetraodon nigroviridis, genomic survey
 sequence.

ACCESSION AL231980.1 GI:7890975
 VERSION AL231980.1
 KEYWORDS GSS: genome survey sequence.
 SOURCE Tetraodon nigroviridis.
 ORGANISM Tetraodon nigroviridis
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
 Tetraodontidae; Tetraodon.

REFERENCE 1 (bases 1 to 1092)
 Roest-Crollius, H., Jallion, O., Dasilva, C., Fitzames, C., Fisher, C.,
 Bouneau, L., Billault, A., Quetier, F., Saurin, W., Bernot, A. and
 Weissenbach, J.
 Characterization and repeat analysis of the compact genome of the
 freshwater pufferfish Tetraodon nigroviridis
 unpublished

TITLE 2 (bases 1 to 1092)
 Roest-Crollius, H., Jallion, O., Dasilva, C., Bouneau, L., Fisher, C.,
 Bernot, A., Fitzames, C., Wincker, P., Brothier, P., Quetier, F.,
 Saurin, W. and Weissenbach, J.
 Human gene number estimate provided by genome wide analysis using
 Tetraodon nigroviridis DNA sequence
 unpublished

JOURNAL 3 (bases 1 to 1092)
 Genoscope.
 Direct Submission
 Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
 TITLE This sequence is a single read and was generated as part of a large
 JOURNAL scale clone-end sequencing project of the Tetraodon nigroviridis
 COMMENT genome. For more information, please take a look at
 http://www.genoscope.cns.fr/Tetraodon.
 FEATURES
 source
 Location/Qualifiers
 1..1092
 /organism="Tetraodon nigroviridis"
 /db_xref="taxon:99883"
 /clone="004C04"
 /clone_lib="G"
 /note="Genoscope sequence ID : C08G004B02XEL-end :
 PUC-01"

BASE COUNT 262 a 294 c 304 g 231 t 1 others
 ORIGIN

Query Match 13.6%; Score 44.4; DB 13; Length 1092;
 Best Local Similarity 58.2%; Pred. No. 0.049; 56; Indels 0; Gaps 0;
 Matches 78; Conservative 0; Mismatches 56; Indels 0; Gaps 0;

OY 194 ccctgagctccaaatgcagctgctggagatggttgcacatcaatgagtgactctga 253
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 52 ccctgtgctggcgaagctgagtgctggcgcagcagctcatcaacatcagatccctga 111
 OY 254 gcagctccagaagaagagcagttccctggtgaagatctcctaagaccctgaagctgg 313
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 112 acggcagcagcagcagcagcctgctctcctcattcaaggagctgttccgcatccgaagcctca 171
 OY 314 taatgagcagctga 327
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 172 cggcagcagcagctga 185

RESULT 14
 A1825748 458 bp MRNA EST 17-DEC-1999
 LOCUS A1825748/c
 DEFINITION wb69608.x1 NCI_CGAP_gc6 Homo sapiens CDNA clone IMAGE:2310903 3',
 mRNA sequence.

ACCESSION A1825748
 VERSION A1825748.1 GI:5446419
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 458)
 NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 Unpublished (1997)
 JOURNAL Contact: Robert Strausberg, Ph.D.
 COMMENT Email: cgaps-remail.nih.gov
 Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
 R. Emmert-Buck, M.D., Ph.D.,
 cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
 Bonaldo, Ph.D.
 cDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 www.bio.llnl.gov/dbfp/image/image.html
 Insert Length: 1112 Std. Error: 0.00
 Seq primer: -400p from Gldco
 High quality sequence stop: 455.
 FEATURES
 source
 Location/Qualifiers
 1..458
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:2310903"
 /clone_lib="NCI_CGAP_gc6"
 /clone_id="NCI_CGAP_gc6"
 /tissue_type="Pooled germ cell tumors"
 /lab_host="DH10B"
 /note="Vector: pTRT-Pac (Pharmacia) with a modified
 polylinker. Site_1: Not I; Site_2: Eco RI; Plasmid DNA
 from the normalized library NCI_CGAP_gc6 was prepared, and
 as circles were made in vitro. Following HAP purification,
 this DNA was used as tracer in a subtractive hybridization
 reaction. The driver was PCR-amplified cDNAs from a pool
 of 5,000 clones made from the same library (clones IDs
 1257096-1258631, 1469064-1470993, and 1475592-1476743).
 Subtraction by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 147 a 95 c 93 g 119 t 4 others
 ORIGIN

Query Match 12.9%; Score 42.2; DB 10; Length 458;
 Best Local Similarity 82.5%; Pred. No. 0.16;
 Matches 47; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

OY 271 gcagttccctggtgaagatcctacaagaccctcagctgtagtcgacgttga 327
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 458 gcagttccctggtgaagatccttncacacccctcagctgtagtcgacgttga 402

RESULT 15
 A1825748 379 bp MRNA EST 25-APR-2001
 LOCUS A1825748/c
 DEFINITION 38503 MARC 2BOV Bos taurus CDNA 5', mRNA sequence.
 ACCESSION A1825748
 VERSION A1825748.1 GI:6680260
 KEYWORDS EST.
 SOURCE COW.
 ORGANISM Bos taurus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 Bovidae; Bovinae; Bos.
 REFERENCE 1 (bases 1 to 379)
 Smith, T.P.L., Grosse, W.M., Freking, B.A., Roberts, A.J., Stone, R.T.,
 Casas, E., Wray, J.E., White, J., Cho, J., Fahrnenkrug, S.C., Bennett,
 G.L., Heaton, M.P., Laegreid, W.A., Rohrer, G.A., Chitko-McKown, C.G.,
 Perera, G., Holt, I., Karamycheva, S., Liang, F., Quackenbush, J. and

Search completed: January 31, 2002, 19:16:11
Job time: 1959 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model
Run on: January 31, 2002, 19:45:37 ; Search time 1599.5 Seconds
(without alignments)
3372.665 Million cell updates/sec

Title: US-09-641-831-3
Perfect score: 327
Sequence: 1 atgatgggaccactgaaga.....ggctggtagtcgcagttga 327

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

1: gb_ba.*

2: gb_hgt.*

3: gb_in.*

4: gb_om.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pl.*

9: gb_pr.*

10: gb_ro.*

11: gb_sts.*

12: gb_sy.*

13: gb_un.*

14: gb_vi.*

15: em_ba.*

16: em_fun.*

17: em_hum.*

18: em_in.*

19: em_om.*

20: em_or.*

21: em_ov.*

22: em_pat.*

23: em_ph.*

24: em_pl.*

25: em_ro.*

26: em_sts.*

27: em_sy.*

28: em_un.*

29: em_vi.*

30: em_htgo_hum.*

31: em_htgo_inv.*

32: em_htgo_rod.*

33: em_htg_hum.*

34: em_htg_inv.*

35: em_htg_rod.*

36: em_htg_other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	327	100.0	327	6	AX088025	Sequence
2	323	98.8	402	6	AX088027	Sequence
3	323	98.8	573	6	AX088023	Sequence
4	209	63.9	6423	10	AF199421	Mus muscu
5	170	52.0	163535	2	AC046189	Homo sapi
6	157.4	48.1	158375	2	AP002859	Homo sapi
7	97.4	29.8	7445	9	HSAPXL	H.sapiens A
8	82.8	25.3	6014	9	AB033028	Homo sapi
9	65	19.9	196542	9	AC002365	Homo sapi
10	45.8	14.0	156413	9	HSAL19E20	Human DNA
11	44.8	13.7	39213	9	AC003047	Homo sapi
12	44.8	13.7	44400	9	AC090481	Homo sapi
c	43.6	13.3	173510	9	AL359272	Human DNA
13	37.8	11.6	2111	9	S39392	Homo sapien
14	37.8	11.6	3984	9	HUMCAP	Human prote
15	36.8	11.3	1485	10	AF002283	Mus muscu
16	36.8	11.3	329709	1	AP002997	Mesorhizo
17	36.8	11.2	3220	9	BC008741	Homo sapi
18	36.6	11.2	3287	9	AF061258	Homo sapi
19	36.2	11.1	1696	10	AF095585	Rattus norv
20	35.8	10.9	1896	10	RNU48247	Homo sapi
21	35.8	10.9	134482	9	AC007335	Homo sapi
22	35.8	10.9	160460	2	AC020766	Homo sapi
c	35.6	10.9	112027	9	AC007006	Homo sapi
23	35.6	10.9	187727	2	AC064806	Homo sapi
c	34.8	10.6	1282	5	SSA308531	Salmo sal
24	34.8	10.6	147197	2	AC091912	Homo sapi
c	34.8	10.6	171280	2	AC024186	Homo sapi
25	34.8	10.6	183349	2	AC027501	Homo sapi
c	34.8	10.6	237508	2	AC009593	Homo sapi
26	34.6	10.6	1567	9	AF345903	Homo sapi
27	34.6	10.6	1669	9	AF345904	Homo sapi
28	34.6	10.6	1686	9	AF345906	Homo sapi
c	34.6	10.6	4542	3	AB036856	Ciona int
29	34.4	10.5	1607	9	HSAL13766	Homo sapi
30	34.2	10.5	1041	10	AB016587	Mus muscu
31	34.2	10.5	1191	10	RNRIT18	R.norvegicu
32	34.2	10.5	1196	10	AB016588	Mus muscu
33	34.2	10.5	1874	10	AB016586	Mus muscu
34	34.2	10.5	3619	9	BC010674	Homo sapi
35	34.2	10.5	3643	9	HUMPTYPH	Human prote
c	34.2	10.5	5121	10	AF074960	Mus muscu
36	34.2	10.5	181850	2	AL354983	Homo sapi
c	34	10.4	968	9	HUMBRP1	Human mRNA
37	34	10.4	4680	1	AF285084S2	Salmonell

ALIGNMENTS

RESULT 1	AX088025	Sequence 3	327 bp	DNA	PAT	17-MAR-2001
LOCUS	AX088025	Sequence 3	327 bp	DNA		
DEFINITION	AX088025	Sequence 3	327 bp	DNA		
ACCESSION	AX088025	Sequence 3	327 bp	DNA		
VERSION	AX088025.1	GI:13396952				
KEYWORDS	human.					
SOURCE	human.					
ORGANISM	Homo sapiens					
REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;					
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.					
TITLE	1 (bases 1 to 327)					
JOURNAL	Turner,C.A., Zambrowicz,B., Friedrich,G., Nehls,M. and Sands,A.T.					
FEATURES	Sequence derived from a human mammary gland cdna library					
source	Patent: WO 0114422-A 3 01-MAR-2001;					
	Lexicon Genetics Incorporated (US)					
	Location/Qualifiers					
	1..327					
	/organism="Homo sapiens"					
	/db_xref="taxon:9606"					

BASE COUNT 91 a 76 c 96 g 64 t
ORIGIN

Query Match 100.0%; Score 327; DB 6; Length 327;
Best Local Similarity 100.0%; Pred. No. 2,8e-88;
Matches 327; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atgataagagaccactgaagacttcccaagcctagtccacattaaacttaacacggcc 60
DB 1 ATGATGAGGACCACTGAAAGACTTCCACAGCCTAGTGCACATTAACCTTAACAGGCC 60
QY 61 accaagaggaaggtacattatcttgaggcatctccctgagaggaagctccctgggtttt 120
DB 61 ACCAAGGGAAGGTACTATTATCTGGAGGCACTCTCGAGGAGAGACTCCCTGGGGTTT 120
QY 121 acttaaaaggttgcttgagcagcagagagaacattatcatcttaaggtcgaaagagg 180
DB 121 ACTCTAAAGGCTGGCTCGAGCAGGAGAACCATTAATCATCTTAAGGTGGAAGAGGG 180
QY 181 ggcagaagcagacacctgagctcccaactgcagcggagatgaggttgccacatcaat 240
DB 181 GGAAGAACACACACCTGAGCTCCAAACTGACGGCTGGGATAGGTTGTGCACATCAAT 240
QY 241 gagtgactctgagcagctccagaagaagagcagttccctggtgaaagatctacaag 300
DB 241 GAGGTGACTCTGAGCAGCTCCAGAAAGAGAGGATTCCCTGTGAAGAGATCTTACAG 300
QY 301 accctcagctggtagtgcgcaagttga 327
DB 301 ACCCTCAGGCTGAGTGCAGCATTTGA 327

RESULT 2
AX088027 402 bp DNA PAT 17-MAR-2001

LOCUS AX088027 Sequence 5 from Patent WO0114422.
ACCESSION AX088027
VERSION AX088027.1 GI:13396953

KEYWORDS

SOURCE human.
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 402)
AUTHORS Turner,C.A., Zambrowicz,B., Friedrich,G., Nehls,M. and Sands,A.T.

TITLE Sequence derived from a human mammary gland cDNA library

JOURNAL Patent: WO 0114422-A 5 01-MAR-2001;
Lexicon Genetics Incorporated (US)

FEATURES
Location/Qualifiers

1..402
/organism="Homo sapiens"
/db_xref="taxon:9606"

BASE COUNT 110 a 91 c 117 g 84 t

ORIGIN

Query Match 98.8%; Score 323; DB 6; Length 402;
Best Local Similarity 100.0%; Pred. No. 4.7e-87;
Matches 323; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atgataagagaccactgaagacttcccaagcctagtccacattaaacttaacacggcc 60
DB 1 ATGATGAGGACCACTGAAAGACTTCCACAGCCTAGTGCACATTAACCTTAACAGGCC 60
QY 61 accaagaggaaggtacattatcttgaggcatctccctgagaggaagcctccctgggtttt 120
DB 61 ACCAAGGGAAGGTACTATTATCTGGAGGCACTCTGAGGAGGAGGAGCTCCCTGGGGTTT 120
QY 121 acttaaaaggttgcttgagcagcagagagaacattatcatcttaaggtcgaaagagg 180
DB 121 ACTCTAAAGGCTGGCTCGAGCAGGAGAACCATTAATCATCTTAAGGTGGAAGAGGG 180

QY 181 ggcagaagcagacacctgagctcccaactgcagcggagatgaggttgccacatcaat 240
DB 181 GGAAGAACACACACCTGAGCTCCAAACTGACGGCTGGGATAGGTTGTGCACATCAAT 240

QY 241 gagtgactctgagcagctccagaagaagagcagttccctggtgaaagatctacaag 300
DB 241 GAGGTGACTCTGAGCAGCTCCAGAAAGAGAGGAGTTCCTGTGGAAGAGATCTTACAG 300

QY 301 accctcagctggtagtgcgcaagttga 323
DB 301 ACCCTCAGGCTGAGTGCAGCATTTGA 323

RESULT 3
AX088023 573 bp DNA PAT 17-MAR-2001

LOCUS AX088023 Sequence 1 from Patent WO0114422.
ACCESSION AX088023
VERSION AX088023.1 GI:13396951

KEYWORDS

SOURCE human.
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 573)
AUTHORS Turner,C.A., Zambrowicz,B., Friedrich,G., Nehls,M. and Sands,A.T.

TITLE Sequence derived from a human mammary gland cDNA library

JOURNAL Patent: WO 0114422-A 1 01-MAR-2001;
Lexicon Genetics Incorporated (US)

FEATURES
Location/Qualifiers

1..573
/organism="Homo sapiens"
/db_xref="taxon:9606"

BASE COUNT 151 a 152 c 152 g 118 t

ORIGIN

Query Match 98.8%; Score 323; DB 6; Length 573;
Best Local Similarity 100.0%; Pred. No. 4.8e-87;
Matches 323; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atgataagagaccactgaagacttcccaagcctagtccacattaaacttaacacggcc 60
DB 1 ATGATGAGGACCACTGAAAGACTTCCACAGCCTAGTGCACATTAACCTTAACAGGCC 60

QY 61 accaagaggaaggtacattatcttgaggcatctccctgagaggaagcctccctgggtttt 120
DB 61 ACCAAGGGAAGGTACTATTATCTGGAGGCACTCTGAGGAGAGAGACTCCCTGGGGTTT 120

QY 121 acttaaaaggttgcttgagcagcagagagaacattatcatcttaaggtcgaaagagg 180
DB 121 ACTCTAAAGGCTGGCTCGAGCAGGAGAACCATTAATCATCTTAAGGTGGAAGAGGG 180

QY 181 ggcagaagcagacacctgagctcccaactgcagcggagatgaggttgccacatcaat 240
DB 181 GGAAGAACACACACCTGAGCTCCAAACTGACGGCTGGGATAGGTTGTGCACATCAAT 240

QY 241 gagtgactctgagcagctccagaagaagagcagttccctggtgaaagatctacaag 300
DB 241 GAGGTGACTCTGAGCAGCTCCAGAAAGAGAGGAGTTCCTGTGGAAGAGATCTTACAG 300

QY 301 accctcagctggtagtgcgcaagttga 323
DB 301 ACCCTCAGGCTGAGTGCAGCATTTGA 323

QY 241 gagtgactctgagcagctccagaagaagagcagttccctggtgaaagatctacaag 300
DB 241 GAGGTGACTCTGAGCAGCTCCAGAAAGAGAGGAGTTCCTGTGGAAGAGATCTTACAG 300

QY 241 gagtgactctgagcagctccagaagaagagcagttccctggtgaaagatctacaag 300
DB 241 GAGGTGACTCTGAGCAGCTCCAGAAAGAGAGGAGTTCCTGTGGAAGAGATCTTACAG 300

QY 241 gagtgactctgagcagctccagaagaagagcagttccctggtgaaagatctacaag 300
DB 241 GAGGTGACTCTGAGCAGCTCCAGAAAGAGAGGAGTTCCTGTGGAAGAGATCTTACAG 300

QY 241 gagtgactctgagcagctccagaagaagagcagttccctggtgaaagatctacaag 300
DB 241 GAGGTGACTCTGAGCAGCTCCAGAAAGAGAGGAGTTCCTGTGGAAGAGATCTTACAG 300

QY 241 gagtgactctgagcagctccagaagaagagcagttccctggtgaaagatctacaag 300
DB 241 GAGGTGACTCTGAGCAGCTCCAGAAAGAGAGGAGTTCCTGTGGAAGAGATCTTACAG 300

QY 241 gagtgactctgagcagctccagaagaagagcagttccctggtgaaagatctacaag 300
DB 241 GAGGTGACTCTGAGCAGCTCCAGAAAGAGAGGAGTTCCTGTGGAAGAGATCTTACAG 300

QY 241 gagtgactctgagcagctccagaagaagagcagttccctggtgaaagatctacaag 300
DB 241 GAGGTGACTCTGAGCAGCTCCAGAAAGAGAGGAGTTCCTGTGGAAGAGATCTTACAG 300

QY 241 gagtgactctgagcagctccagaagaagagcagttccctggtgaaagatctacaag 300
DB 241 GAGGTGACTCTGAGCAGCTCCAGAAAGAGAGGAGTTCCTGTGGAAGAGATCTTACAG 300

QY 241 gagtgactctgagcagctccagaagaagagcagttccctggtgaaagatctacaag 300
DB 241 GAGGTGACTCTGAGCAGCTCCAGAAAGAGAGGAGTTCCTGTGGAAGAGATCTTACAG 300

QY 241 gagtgactctgagcagctccagaagaagagcagttccctggtgaaagatctacaag 300
DB 241 GAGGTGACTCTGAGCAGCTCCAGAAAGAGAGGAGTTCCTGTGGAAGAGATCTTACAG 300

QY 241 gagtgactctgagcagctccagaagaagagcagttccctggtgaaagatctacaag 300
DB 241 GAGGTGACTCTGAGCAGCTCCAGAAAGAGAGGAGTTCCTGTGGAAGAGATCTTACAG 300

QY 241 gagtgactctgagcagctccagaagaagagcagttccctggtgaaagatctacaag 300
DB 241 GAGGTGACTCTGAGCAGCTCCAGAAAGAGAGGAGTTCCTGTGGAAGAGATCTTACAG 300

RESULT 4
AF199421 6423 bp mRNA ROD 13-DEC-1999
LOCUS AF199421 Mms musculus PZ domain actin binding protein Shroom mRNA, complete
DEFINITION
ACCESSION AF199421
VERSION AF199421.1 GI:6467989

[illegible]

```

----- Project Information
Center project name: 19512
Center clone name: 356.M17
----- Summary Statistics
Sequencing vector: M13; W77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 148971 bases at least Q40
Consensus quality: 156340 bases at least Q30
Consensus quality: 159016 bases at least Q20
Insert size: 181000; agarose-fp
Insert size: 160935; sum-of-ctnigs
Quality coverage: 3.6 in Q20 bases; agarose-fp
Quality coverage: 4.0 in Q20 bases; sum-of-ctnigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 27 ctnigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the ctnigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
2096 2195: gap of 100 bp
2196 3520: contig of 1325 bp in length
3521 3620: gap of 100 bp
3621 4777: contig of 1157 bp in length
4778 4877: gap of 100 bp
4878 6660: contig of 1783 bp in length
6661 6760: gap of 100 bp
6761 8530: contig of 1770 bp in length
8531 8630: gap of 100 bp
8631 11149: contig of 2519 bp in length
11150 11249: gap of 100 bp
11250 15298: contig of 4049 bp in length
15299 15398: gap of 100 bp
15399 18533: contig of 3135 bp in length
18534 18633: gap of 100 bp
18634 22044: contig of 3411 bp in length
22045 22144: gap of 100 bp
22145 25578: contig of 3434 bp in length
25579 25678: gap of 100 bp
25679 29571: contig of 3893 bp in length
29572 29671: gap of 100 bp
29672 34405: contig of 4734 bp in length
34406 34505: gap of 100 bp
34506 37415: contig of 2910 bp in length
37416 37515: gap of 100 bp
37516 44162: contig of 6647 bp in length
44163 44262: gap of 100 bp
44263 48837: contig of 4575 bp in length
48838 48937: gap of 100 bp
48938 55115: contig of 6178 bp in length
55116 55215: gap of 100 bp
55216 62194: contig of 6979 bp in length
62195 62294: gap of 100 bp
62295 69547: contig of 7253 bp in length
69548 69647: gap of 100 bp
69648 78050: contig of 8403 bp in length
78051 78150: gap of 100 bp
78151 88041: contig of 9891 bp in length
88042 88141: gap of 100 bp
88142 98560: contig of 10419 bp in length
98561 98660: gap of 100 bp
98661 108146: contig of 9866 bp in length
108147 108246: gap of 100 bp
108247 118855: contig of 10609 bp in length
118856 118955: gap of 100 bp
118956 128543: contig of 9588 bp in length
128544 128643: gap of 100 bp
128644 142647: contig of 14004 bp in length
142648 142747: gap of 100 bp

```

```

FEATURES
source
* 142748 153569: contig of 10822 bp in length
* 153570 153689: gap of 100 bp
* 153670 163535: contig of 9866 bp in length.
Location/Qualifiers
1. 163535
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="4"
/map="4"
/clone="RP11-356M17"
/clone_11D="RPC1-11 Human Male BAC"
1. 2095
/note="assembly-fragment
clone_end:5P6
vector_side:left"
2196. 3520
/note="assembly-fragment"
3621. 4777
/note="assembly-fragment"
4878. 6660
/note="assembly-fragment"
6761. 8530
/note="assembly-fragment"
8631. 11149
/note="assembly-fragment"
11250. 15298
/note="assembly-fragment"
15399. 18533
/note="assembly-fragment"
18634. 22044
/note="assembly-fragment"
22145. 25578
/note="assembly-fragment"
25679. 29571
/note="assembly-fragment"
29672. 34405
/note="assembly-fragment"
34506. 37415
/note="assembly-fragment"
37516. 44162
/note="assembly-fragment"
44263. 48837
/note="assembly-fragment"
48938. 55115
/note="assembly-fragment"
55216. 62194
/note="assembly-fragment"
62295. 69547
/note="assembly-fragment"
69648. 78050
/note="assembly-fragment"
78151. 88041
/note="assembly-fragment"
88142. 98560
/note="assembly-fragment"
98661. 108146
/note="assembly-fragment"
108247. 118855
/note="assembly-fragment"
118956. 128543
/note="assembly-fragment"
128644. 142647
/note="assembly-fragment"
142748. 153569
/note="assembly-fragment"
153670. 163535
/note="assembly-fragment
clone_end:T7
vector_side:right"
BASE COUNT 48438 a 32205 c 31514 g 48770 t 2608 others
ORIGIN

```

Query Match	52.0%	Score 170;	DB 2;	Length 163535;
Best Local Similarity	100.0%;	Pred. No. 1.5e-40;		
Matches	170;	Conservative 0;	Mismatches 0;	Indels 0; Gaps 0;
1	atgatgagaccactgaagactccacaaagcctagtcgcacattaaactcaacacgccc 60			
62590	ATGATGAGGACCACTGAGACTCCACAGCTAGTGCACATTAACCTCAACAGGCC 62639			
61	accaaaggaagggtacatttatctcgaggcattcctcgaggaggagcctccctgggggtttt 120			
62640	ACCAAGGGAAGGTACATTATCTGGAGGCAATTCCTGGAGGAGGAGCTCCCTGGGGGTTTT 62699			
121	actctaagggtggcctggagcagcagacacattaatcatctcaaggt 170			
62700	ACTCTAAGGTTGGCCTGGAGCAGGAGAACCATTAATCATCTCTAAGGT 62749			
RESULT 6				
LOCUS	AP002859	156375 bp	DNA	HTG 29-AUG-2000
DEFINITION	Homo sapiens chromosome 4 clone 225E15 map 4q16-q20, *** SEQUENCING IN PROGRESS ***, 15 unordered pieces.			
ACCESSION	AP002859			
VERSION	AP002859.1	GI:9955382		
KEYWORDS	HTG; HTGS PHASE1.			
SOURCE	Homo sapiens	Homo sapiens DNA, clone:225E15.		
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
AUTHORS	1 (bases 1 to 156375)			
TITLE	Tsal, S. F.			
JOURNAL	Direct Submission			
	Submitted (28-AUG-2000) Shih-Feng Tsal, National Yang-Ming University, Institute of Genetics, 155 Li-Rong St. Section 2, Peitou, Taipei, Taiwan 11221, Republic of China			
	(E-mail: ymptsai@ym.edu.tw, URL: http://genome.ym.edu.tw/, Tel: 886-2-28267043, Fax: 886-2-28264930)			
	These sequences are draft human sequences, not finished sequences. These sequences are unordered pieces. And gaps between the contigs of the same clone are represented as 100 N.			
COMMENT	<p>* NOTE: This is a 'working draft' sequence. It currently consists of 15 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.</p> <p>* This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.</p> <p>* 1 21185: contig of 21185 bp in length</p> <p>* 21186 21285: gap of 100 bp</p> <p>* 21286 22928: contig of 1643 bp in length</p> <p>* 22929 23028: gap of 100 bp</p> <p>* 23029 26285: contig of 3257 bp in length</p> <p>* 26286 26385: gap of 100 bp</p> <p>* 26386 30606: contig of 4221 bp in length</p> <p>* 30607 30706: gap of 100 bp</p> <p>* 30707 36924: contig of 6218 bp in length</p> <p>* 36925 37024: gap of 100 bp</p> <p>* 37025 40953: contig of 3929 bp in length</p> <p>* 40954 41053: gap of 100 bp</p> <p>* 41054 47843: contig of 6790 bp in length</p> <p>* 47844 47943: gap of 100 bp</p> <p>* 47944 57174: contig of 9231 bp in length</p> <p>* 57175 57274: gap of 100 bp</p> <p>* 57275 64994: contig of 7720 bp in length</p> <p>* 64995 65094: gap of 100 bp</p> <p>* 65095 76523: contig of 11429 bp in length</p> <p>* 76524 76623: gap of 100 bp</p> <p>* 76624 88306: contig of 11683 bp in length</p> <p>* 88307 88406: gap of 100 bp</p> <p>* 88407 101162: contig of 12756 bp in length</p> <p>* 101163 101262: gap of 100 bp</p> <p>* 101263 115436: contig of 14174 bp in length</p>			

Fri Feb. 1 08:56:28 2002

RESULT 9

AC002365 196542 bp DNA PRI 22-AUG-1998
 LOCUS Homo sapiens chromosome X clone U177G4, U152H5, U168D5, U17A6,
 DEFINITION U172D6, and U186B3 from Xp22, complete sequence.
 AC002365 U82630 U83511
 VERSION AC002365.1 GI:2358015
 KEYWORDS HTG.
 SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS 1 (bases 1 to 196542)
 Muzny,D., Arenson,A.D., Adams,C., Brundage,E., Bunac,C.,
 Carvelli,K., Chacko,J., Chen,J., Di,W., Ding,Y., Dugan,S.,
 Durbin,J., Forcum,J., Ganesh,R., Garcia,C., Goodman,M.,
 Gorrell,J.H., Haywood,M., Hernandez,J., Jackson,L., Jin,S.,
 Kamal,R., Karpach,S., Kovar,C., Leal,B., Li,Y., Lichtarge,O.,
 Liu,W., Logan,O., Lu,J., Ly,T., Martinez,C., Oswal,G., Perez,L.,
 Rashid,N.D., Rowland,K., Savage,L., Scherer,S.E., Shen,H.,
 Simon,M., Stovall,K., Timms,K.M., Todd,J., Vo,O., Williamson,A.,
 Worley,K.C., Yu,W., Chinault,C., Nelson,D. and Gibbs,R.A.
 Direct Submission
 Unpublished
 2 (bases 1 to 196542)
 Chiu,M.W.

TITLE

JOURNAL Direct Submission
 REFERENCE Submitted (23-JUL-1997) Molecular and Human Genetics, Baylor
 AUTHORS College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
 TITLE 3 (bases 1 to 196542)
 REFERENCE Muzny,D., Ansari-Lari,M.A., Timms,K.M., Yu,W., Dugan,S., Lu,J.,
 AUTHORS Shen,Y., Rowland,K., Liu,W., Perez,L., Ding,Y., Gonzalez,O.,
 Haywood,M., Jain,A., Leal,B., Logan,O., Nguyen,V., Savage,L.,
 Shen,H., Worley,K., Chen,E., Forcum,J., Arenson,A.D., Chiu,M.W.,
 Gorrell,J.H., Brundage,E., Di,W., Chinault,C., Nelson,D. and
 Gibbs,R.A.

TITLE

JOURNAL Direct Submission
 REFERENCE Submitted (20-AUG-1997) Molecular and Human Genetics, Baylor
 AUTHORS College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
 TITLE 4 (bases 1 to 196542)

REFERENCE

AUTHORS Chiu,M.W.
 TITLE Direct Submission
 JOURNAL Submitted (04-SEP-1997) Molecular and Human Genetics, Baylor
 College of Medicine, One Baylor Plaza, Houston, Texas 77030, USA
 REFERENCE 5 (bases 1 to 196542)

REFERENCE

AUTHORS Worley,K.C.
 TITLE Direct Submission
 JOURNAL Submitted (22-AUG-1998) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Sep 4, 1997 this sequence version replaced gi:1766071 gi:1773380
 gi:2289024.

COMMENT

Sequencing is completed to a minimum standard of double strand
 coverage with a minimum of 2 clones and 2 reads with no ambiguities
 or 2 chemistries with a minimum of 2 clones and 3 reads with no
 ambiguities. If the sequence quality does not meet this standard,
 it will be indicated in the annotation.

The repeat regions shown were identified using RepeatMasker by
 Adrian Smit.

Sequence similarities were identified using Powerblast by Jinghui
 Zhang.

Exon/Intron boundaries of identified genes were chosen if there
 were canonical splice junctions that maintained sequence continuity
 across the splice junctions.

FEATURES

source

1. 196542
 /organism="Homo sapiens"
 /db_xref="taxon:9606"

/clone="U177G4"
 /chromosome="X"
 /clone_lib="Livermore Chromosome X Cosmid Library
 LLOXNC01"
 /map="Xp22"
 complement(760..1057)
 /rpt_family="AluJb"
 1853..2194
 /rpt_family="AluSg"
 2305..2384
 /rpt_family="AluSx"
 3418..3630
 /rpt_family="MER20"
 complement(3806..4096)
 /rpt_family="AluSx"
 4138..4537
 /rpt_family="L1MB2"
 complement(4549..4841)
 /rpt_family="AluJo"
 complement(5421..5528)
 /rpt_family="MER58B"
 complement(7001..7132)
 /rpt_family="FAM"
 7773..7903
 /rpt_family="AluJ"
 7910..8215
 /rpt_family="AluYa5"
 8224..8401
 /rpt_family="AluJ"
 complement(8932..9232)
 /rpt_family="AluY"
 complement(9458..9481)
 /rpt_family="POLY_A"
 11579..11618
 /rpt_family="AT-rich"
 complement(11619..11916)
 /rpt_family="AluSg"
 complement(11917..12252)
 /rpt_family="AluJb"
 complement(12328..12633)
 /rpt_family="AluSx"
 complement(12665..13117)
 /rpt_family="L1MB3"
 complement(13236..13267)
 /rpt_family="L1MAL"
 13323..13604
 /rpt_family="AluJb"
 13983..14378
 /note="Similar to sequence with GenBank Accession number
 236840"
 15384..15565
 /rpt_family="MER58B"
 15565..15956
 /rpt_family="AluSg"
 16091..16660
 /rpt_family="L1MB7"
 join(<17023..17176,34350..34481,37731..40071,41563..41663,
 75548..76243,80507..81058,82568..82739,88014..88286,
 90044..>92814)
 /gene="HSAPXL"
 /product="APXL"
 join(<17023..17176,34350..34481,37731..40071,41563..41663,
 75548..76243,80507..81058,82568..82739,88014..88286,
 90044..>92814)
 /gene="HSAPXL"
 >17023..>92814
 /gene="HSAPXL"
 join(<17023..17176,34350..34481,37731..40071,41563..41663,
 75548..76243,80507..81058,82568..82739,88014..88286,
 90044..>9310)
 /gene="HSAPXL"
 /note="human homolog of Xenopus laevis APX gene, similar
 to sequence encoded by GenBank Accession Number X83543"

```

/codon_start=3
/product="APXL"
/protein_id="AAC32592.1"
/db_xref="GI:1773381"
/translation="IEGSKAAAVDKLAGEDEIVGINDIGSGFROEAICLYKSHKT
LKLVAKRKSELGMRPHSHWATKFSDSHPPLASPTSTSGCSWSGRHASSSSDLS
SSWEQNTLQRLDHFSSLSVSLDHSRLVANSSSIDHLSHRSKSDSYGSEFT
SSSTDPHTLSKADTSNAENILYTVGLMEPRGGRQAQADPGSEKELSCFPYV
GDSGGRPREYNAEPKLAPEGNSNGVYVDPKKKAPSPPPPPPLPLKSEFATK
HEKADGSPVSEAAAOHETALAOAORGRGRRPRLDRRSPHPSISLKGSGGCPPO
EADAGSPSPKDGASRLQALSSSDVFRPPSPRSGRPPIYSHPSICAPSLQOE
GAASFONDSPPQVYRGSLSCDOKLGSWQPCFVQDLDLAOLMAGMCPSTALALE
SLPPTVGSPRHHLPQEPGPDPARETGTCTYLDGAECSAGAEPPASVAEKASQ
RLAASITWADGESSRICPQETPLHSLTDEGRKRESSPESDSTRPPPPASAEKASQ
RSDRATITLNEIOMHAKRLQKSRSTVALTAAGEADSGMRAGIGGTGTAGPIAGT
YDHLKEAARVLRATSEFRRLDPPNGLYTESLEHMRGDDDTYHPHEAGLADPSP
STSGPHPRIGRRRTAEOKLSYSEPEKNENGTATGYSYPHOHPHPSPTVTFA
DMKPFETSKYVPOPPAKOALHGIIPRKPRTAGTCTGTEWSTTISLGSLN
AHSAAEKAGTSDLPRLGTFAEYQASWQRKPLEKSSGRCHSADDDIVDSLQOER
PQHVGSRSSPSIDHVKQASVLELRQADGEPREELPSAVRAEEGOSTRQADAO
CREGSPGSOHPPOKAPNPPTFSELHRCAPDELPRGRCAGTLPRDYRSEESTP
ADLGPRAQSGSPPLHARGODSWPVSSALSKRPAPORPPRPRRRYRATDGAADA
PYGVLAGRPPTSPASLDVYVARLSLSHSVPSSAOPQDTPKATVCEPSSQHVADA
SRPLPPLPPEKQHLRLOTATMERSRSPPOPAOKLDPKLLIODESTRIRVM
DNNTVKNVPIKIVHSESOPEKESROSLACPAPLPGLEKDDIKITLSTEQYR
FCLYTRQAGEPAPRPAQAEQPLQTPVPEKDRCTSPGLSYKAKAKTYEDKSE
ELAREIVGKDKSLADLPDSVKIKITMDMEGIFPDEHLEAOQRRLPKRISPR
STEERKEPSPAASLATNSTYTSAPKAEKLLIKMDLQEOHEEDSGDLHDLL
SVKKELTESISRLQVAREARESLIEDVQANTVIGAEYEAIVKGVCKSEPDKRMF
IGDLDKYNVLLSLSLGRLARVENALNLDGNSPDGROSLEKORYLIDQHEDAELK
ENDRERIVFDILANYLSEESLADIEHVFVKMSALLIEQRELEKHIDGEBOLCLL
DSIQPERCK"
/complement(17976..18074)
/repeat_region
/repeat_family="HY3"
18162..18473
/repeat_family="AluUb"
19636..19717
/repeat_family="MER21A"
/complement(19798..19937)
/repeat_region
/repeat_family="LIMB3"
/complement(20181..20327)
/repeat_region
/repeat_family="FRAM"
/complement(22367..22401)
/repeat_region
/repeat_family="At_rich"
22406..22461
/repeat_family="(GGA)n"
/complement(22462..22627)
/repeat_region
/repeat_family="AluUo"
/complement(22638..22933)
/repeat_region
/repeat_family="AluSx"
/complement(26881..26957)
/repeat_region
/repeat_family="MER5A"
/complement(27726..27762)
/repeat_region
/repeat_family="POLY_A"
/complement(27876..27977)
/repeat_region
/repeat_family="AluUo/FRAM"
/complement(28048..28069)
/repeat_region
/repeat_family="At_rich"

```

Query Match 19.9% Score 65; DB 9; Length 196542;
 Best Local Similarity 64.1% Pred No 8.7e-09;
 Matches 98; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

```

QY 175 gaaggggcaaacagaccctgagctccaactgcaggctggggatgaggtgtgcac 234
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 17028 GAAGAGGCGAGTAAAGCCGGGCGGTGACAGTACTGCTGAGATGATGATCGCGG 17087
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 235 atcaatgaggtgagctctgagcagctccagaaagagggcagattccctgtgaaagatcc 294
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 17088 ATCAATGACATTGCTCTCAGGCGTTAGACAGAGCGATTGCTGGTGAAGGGGTCC 17147
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 295 tacaagaccctcagctgtgagtgagcagttga 327
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

DB 17148 CATAGACCCCTGAGCTGCTGTCAAAAGTAA 17180

RESULT 10
 LOCUS HSA119E20
 DEFINITION HSA119E20 156413 bp DNA PRI 14-SEP-2000
 Human DNA sequence from clone RP11-119E20 on chromosome
 Xp11.21-11.23 Contains part of the gene for KIAA1202 protein, ESTs,
 STS and GSSs, complete sequence.

ACCESSION AL121865.7 GI:8979790
 VERSION 1
 KEYWORDS HTG; KIAA1202.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Euteria; Primates; Catarrhini; Homiidae; Homo.
 TITLE Direct Submission
 JOURNAL Submitted (27-Jun-2000) Sanger Centre, Hinxton, Cambridgeshire,
 CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
 requests: clonerequests@sanger.ac.uk

COMMENT
 On Jul 8, 2000 this sequence version replaced gi:8977862.
 During sequence assembly data is compared from overlapping clones.
 Where differences are found these are annotated as variations
 together with a note of the overlapping clone name. Note that the
 variation annotation may not be found in the sequence submission
 corresponding to the overlapping clone, as we submit sequences with
 only a small overlap as described above.

This sequence has been finished according to sequence map criteria
 as follows. An attempt is made to resolve all sequencing problems,
 such as compressions and repeats, but not necessarily within known
 annotated human repeat sequence elements (e.g. Alu). Where the
 sequence is ambiguous, there is an annotation using the 'unsure'
 feature key.

The following abbreviations are used to associate primary accession
 numbers given in the feature table with their source databases:
 Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WormPeP; Information
 on the WormPeP database can be found at
 http://www.sanger.ac.uk/Projects/C.elegans/WormPeP
 was generated from part of bacterial clone contigs of human
 chromosome X, constructed by the Sanger Centre Chromosome X Mapping
 Group. Further information can be found at
 http://www.sanger.ac.uk/HGP/ChrX

RP11-119E20 is from the library RPI1-11.1 constructed at the
 Roswell Park Cancer Institute by the group of Pieter de Jong. For
 further details see http://bacpac.med.bufrato.edu/
 VECTOR: pBAC3.6
 This sequence is the entire insert of clone RP11-119E20.

FEATURES

source
 1..156413
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="X"
 /map="p11.21-11.23"
 /clone="RP11-119E20"
 /clone_11b="RPI1-11.1"
 8..593
 /note="match: GSS: Em:AQ350374"
 1965..2081
 /note="12 repeat: matches 2602..2708 of consensus"
 2374..2464
 /note="12 repeat: matches 2610..2686 of consensus"
 2607..2755
 /note="MIR repeat: matches 27..176 of consensus"
 join(3161..3312,60790..60924,63430..65920,71423..71484,
 90914..91717,96285..96465,100563..100832,102134..103400)
 /gene="DA119E20.1"
 /note="match: CDNs: Em:AB033028 Em:X83543 Em:AB040914"
 /evidence="not_experimental"
 /product="DA119E20.1 (KIAA1202 protein)"
 3161..103400
 /gene="DA119E20.1"
 gene


```

DB      3276 AAGGCTCCTCCGATTCACAGCTGATTCAGAGGATTA 3316
RESULT  11
AC003047 39213 bp DNA PRI 30-MAY-1998
LOCUS    Homo sapiens Xp22 Cosmid U152D7 (Lawrence Livermore human cosmid
DEFINITION library) complete sequence.
ACCESSION AC003047
VERSION   AC003047.1 GI:3169149
KEYWORDS  HTG.
SOURCE    human.
ORGANISM  Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 39213)
AUTHORS   Muzny D., Areson A.D., Adams C., Brundage E., Bunac C.,
           Carvill J., Chacko J., Chen J., Di W., Ding X., Dugan S.,
           Durbin J., Forcum J., Ganes R., Garcia C., Goodman M.,
           Gorrell J.H., Haywood M., Hernandez J., Jackson L., Jin S.,
           Kampal R., Karpathy S., Kovar C., Leal B., Li Y., Lichtarge O.,
           Liu W., Logan O., Lu J., Ly T., Martinez C., Oswal G., Perez L.,
           Rashid N.D., Rowland K., Savage L., Scherer S.E., Shen H.,
           Simon M., Stovall K., Timms K.M., Todd J., Vo O., Williamson A.,
           Worley K.C., Yu W., Chinault C., Nelson D. and Gibbs R.A.
TITLE      Direct Submission
JOURNAL    Unpublished
REFERENCE 2 (bases 1 to 39213)
AUTHORS   Morley K.C.
TITLE      Direct Submission
JOURNAL    Submitted (03-NOV-1997) Molecular and Human Genetics, Baylor
           College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3 (bases 1 to 39213)
AUTHORS   Morley K.C.
TITLE      Direct Submission
JOURNAL    Submitted (30-MAY-1998) Human Genome Sequencing Center, Department
           of Molecular and Human Genetics, Baylor College of Medicine, One
           Baylor Plaza, Houston, TX 77030, USA
COMMENT    On May 30, 1998 this sequence version replaced gi:3108011.
           Sequencing is completed to a minimum standard of double strand
           coverage with a minimum of 2 clones and 2 reads with no ambiguities
           or 2 chemistries with a minimum of 2 clones and 3 reads with no
           ambiguities. If the sequence quality does not meet this standard,
           it will be indicated in the annotation.
           The repeat regions shown were identified using RepeatMasker by
           Adrian Smith.
           Sequence similarities were identified using Powerblast by Jinghui
           Zhang.
FEATURES             source
   Exon/Intron boundaries of identified genes were chosen if there
   were canonical splice junctions that maintained sequence continuity
   across the splice junctions.
   Location/Qualifiers
       1..39213
           /organism="Homo sapiens"
           /db_xref="taxon:9606"
           /clone="U152D7"
           /chromosome="X"
           /clone_lib="Lawrence Livermore human cosmid library"
           /map="Xp22"
           complement(9..118)
           /rpt_family="MLT1A1"
           119..438
           /rpt_family="L1MB6"
           complement(447..760)
           /rpt_family="AluYo"
           repeat_region(complement(968..1179)
           /rpt_family="MIR"
           1308..1602
           /rpt_family="AluYb"
           repeat_region

```

```

repeat_region complement(1710..1835)
/rpt_family="FLAM_A"
repeat_region complement(2789..3087)
/rpt_family="AluSx"
repeat_region complement(3095..3126)
/rpt_family="AT_rich"
repeat_region complement(3245..3375)
/rpt_family="MER5A"
repeat_region 3622..3712
/rpt_family="(GGA)n"
repeat_region complement(3762..3788)
/rpt_family="AT_rich"
repeat_region complement(3789..4087)
/rpt_family="AluSx"
repeat_region 4514..4801
/rpt_family="AluSx"
repeat_region 4804..5096
/rpt_family="AluSx"
repeat_region complement(5643..5927)
/rpt_family="AluYo"
repeat_region 5928..6181
/rpt_family="AluYo"
repeat_region 6505..6701
/rpt_family="MER5B"
repeat_region 6715..6995
/rpt_family="AluYb"
repeat_region 7006..7092
/rpt_family="MER5B"
repeat_region 7191..7227
/rpt_family="(CA)n"
repeat_region 7428..7724
/rpt_family="AluSx"
repeat_region 8196..8266
/note="low coverage"
repeat_region complement(8304..8355)
/rpt_family="GC_rich"
repeat_region 8705..8789
/rpt_family="(GGA)n"
repeat_region 9136..9429
/rpt_family="AluSx"
repeat_region 9430..9580
/rpt_family="AluYb"
repeat_region 9588..9625
/rpt_family="(CA)n"
repeat_region 9627..9663
/rpt_family="AT_rich"
repeat_region 10920..11037
/rpt_family="AluYb"
repeat_region 11146..11196
/note="low coverage"
repeat_region 11203..11335
/rpt_family="FRAM"
repeat_region complement(11351..11652)
/rpt_family="AluSx"
repeat_region complement(12228..12254)
/rpt_family="AT_rich"
repeat_region complement(12257..12556)
/rpt_family="AluSx"
repeat_region 12557..12590
/rpt_family="AT_rich"
repeat_region 12704..12994
/rpt_family="AluYo"
repeat_region 12995..13020
/rpt_family="(GA)n"
repeat_region 13023..13093
/rpt_family="MER5A"
repeat_region 13667..13854
/rpt_family="AluSx"
repeat_region complement(14127..14233)
/rpt_family="MER5A"
repeat_region complement(14368..14670)
/rpt_family="AluSx"
repeat_region 15204..15494

```

[illegible]

flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

QUALSTAT-REPORT-----

----- Summary Statistics -----

Contig Length: 44400
Phrap values in estimate: 44245
Average error rate (BCW-Phrap estimate): 7.63605e-07
Fraction of Phrap values less than 40 : 0.00101706
Number of consensus changing edits: 10
Number of N's in consensus : 0

Position	Original+Context	Edited+Context
12913	tcaattctat(n)lalttgatt	tcaattctat(t)lalttgatt
20039	ccattccaat(n)lgaatgatg	ccattccaat(t)lgaatgatg
38071	ggaggaattg(n)lgtcccttac	ggaggaattg(t)lgtcccttac
38096	tgatgatgag(n)atttggtcg	tgatgatgag(t)atttggtcg
44395	catctctg(n)nnnnngcgcc	catctctg(g)nnnnngcgcc
44396	atctctg(n)nnnnngcgcc	atctctg(g)nnnnngcgcc
44397	tctctg(n)nnnnngcgcc	tctctg(g)nnnnngcgcc
44398	ctctctg(n)nnnnngcgcc	ctctctg(g)nnnnngcgcc
44399	ctctg(n)nnnnngcgcc	ctctg(g)nnnnngcgcc
44400	ctgtg(n)nnnnngcgcc	ctgtg(g)nnnnngcgcc

----- Distribution of Quality < 40 Bases -----

Phrap Value Range	Count
5	501
10	451
15	401
20	351
25	301
30	251
35	201
40	151
45	101
50	51
55	01

FEATURES

SOURCE

Version: 1.01 gxf.

Location/Qualifiers

1. 44400

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="Xp"

/clone="RP11-8H7"

/complement(1..2004)

/note="Overlaps bases 2004..1 of clone AC003036"

/function="Overlaps with adjacent clone AC003036"

misc_feature

repeat_region

/rpt_family="MLTJ"

repeat_region

/rpt_family="AluSq"

repeat_region

/rpt_family="MLTJ"

repeat_region

/rpt_family="LMB7"

repeat_region

/rpt_family="MLTJ"

repeat_region

/rpt_family="LMB7"

repeat_region

/rpt_family="MLTJ"

repeat_region

/rpt_family="LMB7"

repeat_region

/rpt_family="MLTJ"

repeat_region

/rpt_family="LMB7"

repeat_region

/rpt_family="MLTJ"

repeat_region

/rpt_family="LMB7"

repeat_region

/rpt_family="MLTJ"

repeat_region

/rpt_family="LMB7"

repeat_region

/rpt_family="MLTJ"

repeat_region

/rpt_family="LMB7"

repeat_region

/rpt_family="MLTJ"

repeat_region

/rpt_family="LMB7"

repeat_region

/rpt_family="MLTJ"

repeat_region

/rpt_family="LMB7"

repeat_region

/rpt_family="MLTJ"

repeat_region

/rpt_family="LMB7"

repeat_region

/rpt_family="MLTJ"

repeat_region

/rpt_family="LMB7"

repeat_region

/rpt_family="MLTJ"

repeat_region

/rpt_family="LMB7"

repeat_region

/rpt_family="MLTJ"

repeat_region

/rpt_family="LMB7"

repeat_region

/rpt_family="MLTJ"

repeat_region

/rpt_family="LMB7"

LOCUS AL359272 173510 bp DNA 12-OCT-2000
DEFINITION Human DNA sequence from clone RP11-554P16 on chromosome X. Contains the first coding exon of the gene KIAA1202, STSs, GSSs and a CpG island, complete sequence.
ACCESSION AL359272
VERSION AL359272.9 GI:9650579
KEYWORDS HTG; CpG island; KIAA1202.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS 1 (bases 1 to 173510)
TITLE Bird.C.
JOURNAL Direct Submission
COMMENT Submitted (28-SEP-2000) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
 requests: clonerequest@sanger.ac.uk
 On Aug 1, 2000 this sequence version replaced gi:9588592.
 During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
 This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.
 The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em., EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information on the WORMPEP database can be found at <http://www.sanger.ac.uk/Projects/C/elegans/wormpep> This sequence was generated from part of bacterial clone contigs of human chromosome X, constructed by the Sanger Centre Chromosome X Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/ChrX>
 RP11-554P16 is from the library RPC1-11.2 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see <http://bacpac.med.buffalo.edu/VECTOR: pbACe3.6>
 This sequence is the entire insert of clone RP11-554P16.

FEATURES
 source
 1..173510
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="X"
 /clone="RP11-554P16"
 /clone_lib="RPC1-11.2"
 2..192
 /note="L1MEC repeat: matches 1913. .2098 of consensus"
 7..674
 /note="match: GSS: Em:AQ373815"
 195..511
 /note="AluSp repeat: matches 1. .312 of consensus"
 543..845
 /note="AluSx repeat: matches 1. .304 of consensus"
 950..1127
 /note="L1MEC repeat: matches 1968. .2136 of consensus"
 1190..2306
 /note="L1M4b repeat: matches -180. .888 of consensus"
 2425..3343
 /note="L1M4 repeat: matches 3078. .4092 of consensus"
 3344..3717
 /note="MSTD repeat: matches 1. .394 of consensus"
 3718..4412
 /note="L1M4 repeat: matches 4092. .4274 of consensus"
 4413..4452
 /note="THE1B repeat: matches 1. .42 of consensus"
 4453..4763
 /note="AluSx repeat: matches 1. .310 of consensus"

4764..5111
 /note="THE1B repeat: matches 42. .364 of consensus"
 5112..5774
 /note="L1M4 repeat: matches 4274. .5344 of consensus"
 5775..6080
 /note="AluSg repeat: matches 1. .307 of consensus"
 6081..6296
 /note="L1M4 repeat: matches 5344. .5581 of consensus"
 6299..6923
 /note="L1PA10 repeat: matches 5525. .6165 of consensus"
 7394..7652
 /note="L1M4 repeat: matches 2120. .2383 of consensus"
 7652..10378
 /note="L1PA4 repeat: matches 3414. .6142 of consensus"
 10369..10878
 /note="L1M4 repeat: matches 2350. .2879 of consensus"
 10875..11250
 /note="L1 repeat: matches 4741. .5131 of consensus"
 11247..111619
 /note="L1MA7 repeat: matches 5903. .6284 of consensus"
 11642..11764
 /note="L1M4 repeat: matches 5351. .5475 of consensus"
 11764..11943
 /note="L1 repeat: matches 3846. .4036 of consensus"
 12269..113359
 /note="L1MC3 repeat: matches 6659. .7739 of consensus"
 complement(13126..13490)
 /note="match: STS: Em:HS308ZG9"
 13430..13497
 /note="L1M4 repeat: matches 1995. .2062 of consensus"
 13498..13803
 /note="AluJo repeat: matches 3. .307 of consensus"
 13804..15085
 /note="L1M4 repeat: matches 720. .1995 of consensus"
 15191..15376
 /note="L1M4 repeat: matches 408. .602 of consensus"
 15585..16035
 /note="L1M4 repeat: matches -252. .225 of consensus"
 16058..16114
 /note="L12 repeat: matches 2686. .2745 of consensus"
 16128..16369
 /note="L1MA10 repeat: matches 5928. .6318 of consensus"
 17285..17533
 /note="match: GSS: Em:AQ134567"
 17455..17899
 /note="match: GSS: Em:AQ142071"
 17875..18018
 /note="MIR repeat: matches 96. .252 of consensus"
 18979..19286
 /note="AluDb repeat: matches 2. .308 of consensus"
 19391..19823
 /note="L12 repeat: matches 2284. .2708 of consensus"
 20320..20610
 /note="AluSx repeat: matches 1. .289 of consensus"
 21916..22121
 /note="match: STS: Em:G03829"
 23553..23862
 /note="MER33 repeat: matches 1. .307 of consensus"
 23946..24118
 /note="MIR repeat: matches 77. .259 of consensus"
 24356..24475
 /note="L12 repeat: matches 2624. .2750 of consensus"
 24929..25125
 /note="MIR repeat: matches 23. .256 of consensus"
 25251..25386
 /note="MER63 repeat: matches 2. .768 of consensus"
 25604..25656
 /note="L12 repeat: matches 2657. .2709 of consensus"
 26047..27589
 /note="L1MB8 repeat: matches 4591. .6159 of consensus"
 27590..27886
 /note="AluSx repeat: matches 1. .297 of consensus"
 27887..28214

```

repeat_region /note="LIMB8 repeat: matches 4271. .4591 of consensus"
28258. .29096 /note="LIMB8 repeat: matches 3264. .4157 of consensus"
repeat_region /note="LIMB8 repeat: matches 1. .383 of consensus"
29118. .29521 /note="LIMB8 repeat: matches 2496. .3231 of consensus"
repeat_region /note="LIMB8 repeat: matches 2. .299 of consensus"
30382. .30685 /note="LIMB8 repeat: matches 2110. .2496 of consensus"
repeat_region /note="LIMB8 repeat: matches 5820. .6156 of consensus"
30686. .31069 /note="LIMB8 repeat: matches 5374. .5552 of consensus"
repeat_region /note="LIMB8 repeat: matches 1. .294 of consensus"
31910. .31976 /note="LIMB8 repeat: matches 1951. .2330 of consensus"
repeat_region /note="LIMB8 repeat: matches 969. .2197 of consensus"
31978. .32359 /note="LIMB8 repeat: matches 1. .364 of consensus"
repeat_region /note="LIMB8 repeat: matches 2576. .2710 of consensus"
33354. .33592 /note="LIMB8 repeat: matches 194. .257 of consensus"
repeat_region /note="LIMB8 repeat: matches 194. .257 of consensus"
33935. .34790 /note="LIMB8 repeat: matches 194. .257 of consensus"
repeat_region /note="LIMB8 repeat: matches 194. .257 of consensus"
34856. .35219 /note="LIMB8 repeat: matches 194. .257 of consensus"
repeat_region /note="LIMB8 repeat: matches 194. .257 of consensus"
36043. .36159 /note="LIMB8 repeat: matches 194. .257 of consensus"
repeat_region /note="LIMB8 repeat: matches 194. .257 of consensus"
38455. .39057 /note="LIMB8 repeat: matches 194. .257 of consensus"
repeat_region /note="LIMB8 repeat: matches 194. .257 of consensus"
41978. .42373 /note="LIMB8 repeat: matches 194. .257 of consensus"
repeat_region /note="LIMB8 repeat: matches 194. .257 of consensus"
42056. .42087 /note="LIMB8 repeat: matches 194. .257 of consensus"
repeat_region /note="LIMB8 repeat: matches 194. .257 of consensus"
42311. .42540 /note="LIMB8 repeat: matches 194. .257 of consensus"
repeat_region /note="LIMB8 repeat: matches 194. .257 of consensus"
42824. .42914 /note="LIMB8 repeat: matches 194. .257 of consensus"
repeat_region /note="LIMB8 repeat: matches 194. .257 of consensus"
43119. .43173 /note="LIMB8 repeat: matches 194. .257 of consensus"
repeat_region /note="LIMB8 repeat: matches 194. .257 of consensus"
43122. .43185 /note="LIMB8 repeat: matches 194. .257 of consensus"
repeat_region /note="LIMB8 repeat: matches 194. .257 of consensus"
44700. .44743 /note="LIMB8 repeat: matches 194. .257 of consensus"
repeat_region /note="LIMB8 repeat: matches 194. .257 of consensus"
45956. .46137 /note="LIMB8 repeat: matches 194. .257 of consensus"
repeat_region /note="LIMB8 repeat: matches 194. .257 of consensus"
46401. .46617 /note="LIMB8 repeat: matches 194. .257 of consensus"
repeat_region /note="LIMB8 repeat: matches 194. .257 of consensus"
46729. .47035 /note="LIMB8 repeat: matches 194. .257 of consensus"
repeat_region /note="LIMB8 repeat: matches 194. .257 of consensus"
47439. .47682 /note="LIMB8 repeat: matches 194. .257 of consensus"
repeat_region /note="LIMB8 repeat: matches 194. .257 of consensus"
48430. .48527 /note="LIMB8 repeat: matches 194. .257 of consensus"

Query Match 13.3%; Score 43.6; DB 9; Length 173510;
Best Local Similarity 67.8%; Pred. No. 0.026;
Matches 61; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

QY 94 ctggaaggagagcctccctgggttttactctaaagggtgctgagacaggaacaa 153
DB 84025 CTCGAAGGGGGGACCCCTTCAAGGGGGGCTGGAACACTGTGAGCG 83966

QY 154 ttaatcatctcctaagtcgaaggaaggagc 183
DB 83965 CTCACAGTGTCTTAAGTAAGAACTGGCGGC 83936

RESULT 14
LOCUS S39392 2111 bp mRNA PRI 05-MAR-2001
DEFINITION Homo sapiens protein tyrosine phosphatase mRNA, partial cds.
ACCESSION S39392

```

```

VERSION S39392.1 GI:250890
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 2111)
AUTHORS Arimura,Y., Hinoda,Y., Itoh,F., Takekawa,M., Tsujisaki,M.,
Adachi,M., Inai,K. and Yachi,A.
TITLE cDNA cloning of new protein tyrosine phosphatases in the human
JOURNAL Tumor Biol. 13 (3), 180-186 (1992)
MEDLINE 92327504
REMARK Genbank staff at the National Library of Medicine created this
entry [NCBI gidsq 108079] from the original journal article.
This sequence comes from Fig. 3.
FEATURES
Source location/Qualifiers
1..2111
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="C16"
/tissue-type="colon"
<1..>2111
/note="Primer: This sequence comes from Fig. 3"
/codon_start=3
/product="protein tyrosine phosphatase"
/protein_id="AAB22439.2"
/db_xref="GI:13195754"
/translation="OHSGIKOSEAESCTINIAFLDFYGVLEHSGRDLNMLNIGIA
SAGVAAYRYCTSPFYPMVNTIKTSFKRKFTHOROKAQRSHIYAFNLRNRSK
NMKSCVGHHTFPQAKKILPOKKNVLSQYWMGSRNKKSYNNQYCKKVGGMWMA
MRSLSVHETKSLPSPSPPTTPMWRSPRIHRIRKRRHSADNLNENYITETED
VFYIKGSLAPDSDSEVSONRSPHOSLSNNPAQSTLYTKSSVSPSSNAGSCS
PDGVDOQLDHFHYTKGSTEDASQYCDKNDGDSYLVLRITPEDGKFGNLRG
GVDMKPLVVRIRNPSPADICIRLNGDQIVLNGDISEHTHDVYVPIKASRS
HSRELATVIRRAVRSFADPKSEDELNPPEALFPMCPGEGDTLEGSMOLKASRS
GTVLIQPOLYRKRYGALITPAKLPONLDKRYRDVLPYDTLVILGQNEPIYASVY
MTEIPLAHLVWKYLAATGSLPHTCAGQAFQVYWDOKLSIYVLTTLTERGRRCKQYMP
DPPVYMNHGPHICQSEDCITAYSRSMULTNTNQTGEHVTYHLYQYAMPDRIPTD
SSDFEAFVYRSLVDSPEVYVHCSAIGRTGVLVYMETAMCLTERNLPYPIQIVR
KMDQRAAMVQTSQYKRCVE"
BASE COUNT 594 a 528 c 510 g 479 t
ORIGIN

Query Match 11.6%; Score 37.8; DB 9; Length 2111;
Best Local Similarity 53.8%; Pred. No. 1.1;
Matches 78; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

QY 97 gaaggagagagcctccctgggttttactctaaagggtgctgagacaggaacaa 156
DB 969 GATGAAGATGGAATAATTGGATTTAATCTTAAGGAGAGTGATCAAAAGATGCTCT 1028

QY 157 atcaatcctaagtcgaaggaagggaagacacacctgaagctcaactgaagct 216
DB 1029 GTGGTATCAAGGATTAACCCAGAGTCACCTCGCGACCTGCATTCCTTAAGTGAACGA 1088

QY 217 ggggatgaggttgctgacataag 241
DB 1089 GGGGATCAAAATGCTGTTAATCAATG 1113

RESULT 15
LOCUS HUMCAP 3984 bp mRNA PRI 27-APR-1993
DEFINITION Human protein tyrosine phosphatase mRNA, complete cds.
ACCESSION M64572
VERSION M64572.1 GI:179912
KEYWORDS tyrosine phosphatase.
SOURCE Homo sapiens cDNA to mRNA.
ORGANISM Homo sapiens
DEFINITION Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

```

REFERENCE	AUTHORS	TITLE	JOURNAL
1 (bases 1 to 3984)	Yang, Q. C. and Tonks, N. K.	Isolation of a cDNA clone encoding a human protein-tyrosine phosphatase with homology to the cytoskeletal-associated proteins band 4.1, ezrin, and talin	Proc. Natl. Acad. Sci. U.S.A. 88, 5949-5953 (1991)

```

Location/Qualifiers
1..3984
  organism="Homo sapiens"
  db_xref="taxon:9606"
  cell_type="HELA CELLS"
24..2765
  note="putative"
/codon_start=1
/product="protein-tyrosine phosphatase"
/protein_id="AAA35647.1"
/db_xref="gi:179913"
/translation="MTSLRALGGRINIRTSELPKEKTRSEVICSIHFLDGVGVQTFK
LTKRGVGLDMVHNLGVLTKEFYGLQHDQDDSDSPRWLEASRP IKRLQAGFPCT
LVFRVETIPDPTLQQQTRILYFLQLKMGDCEGRLLCTPLNSAVVLASVAVQSHFGD
YNSSHHEGYLSDSHFIPDQNEDEFLTKESLHEQSHGLKQSAEACY INIATLDYFG
VELHSGRDLDMIGTASAGVAYRYKICTSFYPMVNLIKTSFKRKFETHOROKY
AESRSHVAFNMILNTRGKSNLWKSQVEHHFTFOAKLLPQENLVSQYWTMSRNTKK
SYNNQCKYKIGVMYMNAMRSLSVHEHLETKSLSPSRQITPNWRSPRLRIEIRKAP
HSSADLANEMTYITETEDFYTYKGLAPQSDSVSPQSPRSELSNNPQAQSL
TOKSSSVSPNSAGSCSPDGVQQLDDFHRTVKSGTDEASQYCYCDKNDNGDSYL
VLIRTPDDEGKFGFKGVQDKWPLVYSRINPSPADTCIPKLNQEQIVLNGRD
ISEHTHDQVMFKKASRSHSRELAIVIRRAVRSFADFKSEDELNQLPEALTFMCP
EGDTHLEGSMAQLKGESGTVLIOEQLYRKKPGLAITFAKLPDLNKNRYKDVLPY
DTRVLQGNEDY INASVYMEIIPAAANLKNKYITATQGLPHCAQFQWVVDOKLSLI
VMLTITRGRFKCHQWPDVPMVNHGFGHQCEDECTIAIVSRMLVNTQTGEEA
HVTHLQYVAMPDHGIPDSDSDFEFTVYRSLRVDSPVHCSAGIGRTGVLTVEE
TAMCLTERRLPTIPLDIVRKMQRQAMVQTSQYKRFVCEAILRVYEEGLVQMLDPS"
954 c 930 g 1014 t
1086 a

```

	Query Match	11.6%;	Score 37.8;	DB 9;	Length 3984;
	Best Local Similarity	53.8%;	Pred. No. 1.1;		
	Matches	78;	Conservative	0;	Mismatches 67; Indels 0; Gaps 0;
Qy	97	gagggaaggagctccctgggttttacttaagggtggcctggagcagcgagaccatta	156		
Db	1569	GATCAAGATGGAATAATTGGATTATCTTAAGGGAGGTGGATCAAAAGATGCCCTCTT	1628		
Qy	157	atcatctctaaggtcgaagggggaagcagacagcacctcctgagctccaaactcgaggct	216		
Db	1629	GTGCTATCAAGGATAACCCAGAGTCACCTCGGCACACCTTCATTCTTAAGCTGAACGAA	1688		
Qy	217	gggggatgaggtgtgcatcaatg	241		
Db	1689	GGGGATCAAAATCGTGTTAATCAATG	1713		

Search completed: January 31, 2002, 19:46:38
Job time: 1816 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model
Run on: January 31, 2002, 19:16:11 ; Search time 1929.4 Seconds
(without alignments)
2238.936 Million cell updates/sec

Title: US-09-641-831-5
Perfect score: 402
Sequence: 1 atgatggagaccactga.....atcatgcaactgcagcttga 402

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 537289281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- EST:*
1: em_estfun:*
2: em_esthum:*
3: em_estom:*
4: em_estom:*
5: em_estpl:*
6: em_estba:*
7: em_estov:*
8: em_estov:*
9: em_hic:*
10: gb_estl:*
11: gb_est2:*
12: gb_hic:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rod:*
20: em_gss_vit:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	151.4	37.7	649	11	BG563384 602582446
2	100.8	25.1	1121	12	AK018481 Mus muscu
3	91.2	22.7	321	10	BB504666 BB504666
4	85.8	21.3	481	11	BF463332 UI-M-CGOp
5	84.8	21.1	484	10	AU170966 AU170966
6	82.6	20.5	231	10	AV368091 AV368091
7	76.4	19.0	597	11	BF057621 7x46h06.x
8	72.2	18.0	558	11	BF434691 7p03b06.x
9	64.6	16.1	383	13	AQ244687 HS_2045_A
10	63.8	15.9	354	10	AW001038 w-91h03.x
11	60.2	15.0	309	13	AQ393528 CITBI-E1
12	60.2	15.0	543	13	AQ457616 HS_5097_B

C 13	60.2	15.0	560	13	AQ886165
C 14	59.2	14.7	875	11	BG432009
C 15	58.8	14.6	463	10	AI687027
C 16	57.6	14.3	372	10	BE763418
C 17	57.4	14.3	941	11	BG284962
C 18	57.2	14.2	265	10	AW166706
C 19	57.2	14.2	270	11	BG117722
C 20	57.2	14.2	325	10	AI885995
C 21	57.2	14.2	723	11	BF572167
C 22	57.2	14.2	540	11	AQ769741
C 23	56.8	14.1	404	11	R22776
C 24	56.8	13.9	860	13	AQ743955
C 25	55.8	13.9	824	10	AL042013
C 26	55.6	13.8	527	13	AQ313163
C 27	55.4	13.8	417	10	AI248313
C 28	55.4	13.8	457	10	AI248328
C 29	55.4	13.8	460	10	AI023290
C 30	55.4	13.8	495	11	BG231844
C 31	55.4	13.8	527	13	AQ494912
C 32	55.4	13.8	615	13	AQ768789
C 33	54.6	13.6	596	13	AQ569045
C 34	54.4	13.5	435	10	AI952612
C 35	54.2	13.5	441	13	AQ581530
C 36	54.2	13.5	499	13	AQ581533
C 37	54	13.4	389	10	AA748026
C 38	54	13.4	679	10	AA715049
C 39	54	13.4	932	13	CNS03SP4
C 40	53.8	13.4	214	13	AQ027606
C 41	53.8	13.4	307	13	AQ217481
C 42	53.8	13.4	513	13	AQ206820
C 43	53.8	13.4	583	13	AQ726346
C 44	53.8	13.4	593	10	BE153075
C 45	53.8	13.4	646	13	AQ583051

ALIGNMENTS

RESULT 1
BG563384 649 bp mRNA EST 10-APR-2001
LOCUS 602582446F1 NIH_MGC_76 Homo sapiens cDNA clone IMAGE:4710299 5',
DEFINITION mRNA sequence.
ACCESSION BG563384.1 GI:13571036
VERSION BG563384.1
KEYWORDS EST.
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 649)
AUTHORS NIH-MGC http://mgc.ncbi.nlm.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Title: Procurement: CLONTECH Laboratories, Inc.
CDNA Library Preparation: CLONTECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LUCM1548 row: 1 column: 12
High quality sequence stop: 548.
Location/Qualifiers
1. 649
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4710299"
/clone_lib="NIH_MGC_76"
/lab_host="DH10B (TI phage-resistant)"
/note="Organ: liver; Vector: pDNR-LIB (Clontech); Site:1."


```
lung"
/sex="mixed"
/tissue_type="lung"
/dev_stage="16 days embryo"
/lab_host="DH10B"
/notes="Site_1: Sali; Site_2: BamHI: cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN, Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5',
GAGAGAGAGAGATCCAGACGCTTTTCTTTTCTTTTCTTNN 3'], cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. cDNA went through one round of normalization
to Rot = 10.0 and subtraction to Rot = 185.0. Second
strand cDNA was prepared with the primer adapter of
sequence [5' GAGAGAGAGATTCGAGTTAATAATTATCCCCCCCCCCCC
3']. cDNA was cloned into the XhoI and BamHI sites.
Vector: a modified pBluescript KS(+) after bulk excision
from Lambda FLC I. Cloning sites, 5' end: Sali; 3' end:
BamHI."
```

```

BASE COUNT      52 a      56 c      52 g      71 t
ORIGIN
Query Match      20.5%; Score 82.6; DB 10; Length 231;
Best Local Similarity 71.2%; Fred. No. 3.2e-12;
Matches 109; Conservative 0; Mismatches 44; Indels 0; Gaps
0;

QY      18 agacttccacaagctagtgcacattaaactctaacacggccaccgaaggaaggtacat 77
Db      1 AAAATTGTAGCAGCTAGCCCTACTCTCAAAACCCACATACGACCCCCACCCGAGAGATTCT 60

QY      78 ttatctggaggcattcctggaggagagctccctgggggtttactctaaagggtggcct 137
Db      61 TTATATGGAACCCCTCTCTAGAGGGAGGGGATCCCTGGGCCCTCACCCCTGAAAGGTGGCCT 120

QY      138 ggagcagcgagaaccattaatcatctctaaagt 170
Db      121 GGACGCTGGGAACACCGTTATTCATTTCCAAGGT 153

```

```

RESULT 7
BF057621
LOCUS
DEFINITION
16-OCT-2000
EST
mrna
597 bp
BF057621
744606.x1 NCI_CGAP_OV18 Homo sapiens cDNA clone IMAGE:3478523 3'
similar to SW:APXL_HUMAN Q13796 APICAL-LIKE PROTEIN ;contains
MER22_b2 TAR1 repetitive element ;, mRNA sequence.
BF057621
BF057621.1 GI:10811517
EST.
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 597)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgaps-femail.nih.gov
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
R. Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: M. Bento
Soares, Ph.D. cDNA Library Arrayed by: Christa Prange, The
I.M.A.G.E. Consortium DNA Sequencing by: Washington University
Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL, send email to:
info@image.llnl.gov
Seq primer: -40up from Glibco

```

```

High quality sequence stop: 511.
Location/Qualifiers
  1. 597
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /clone="IMAGE:3478523"
    /clone_lib="NCI_CGAP_Ov18"
    /tissue_type="fibrotheoma"
    /lab_host="DH10B (phage-resistant)"
    /note="organ: ovary; Vector: pT7T3D-Pac (Pharmacia) with a
    modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
    strand cDNA was primed with a Not I - oligo(dT) primer [5',
    TGTTCACATCTGAGTCGAGCGCGCGACATTTTTTTTTTTT 3'];
    double-stranded cDNA was ligated to Eco RI adaptors
    (Pharmacia), digested with Not I and cloned into the Not
    I and Eco RI sites of the modified pT7T3 vector. Library
    went through one round of normalization, and was
    constructed by Bento Soares and M. Fatima Bonaldo.
    125 a 205 c 158 g 108 t 1 others

```

	Query Match	19.0%;	Score 76.4;	DB 11;	Length 597;
	Best Local Similarity	62.2%;	Pred. No. 1.9e-10;		
	Matches 138;	Conservative	0; Mismatches 81;	Indels 3;	Gaps 1;
QY	131	gtgccttgagcagcggaaccattaatcatctctaaggtcgagaagggggcaacgacg	190		
Dd	1	CGGCGCGAGCAGCGGGAGCCCTGTGTCATCACCAAGATTGAAGAGGGCAGTAAAGC--	58		
QY	191	acaacctgagtccaaactgcagccttgggatgaggttgtgcatacataatgaggtgactc	250		
Dd	59	-CGGCGCGGTGCACAAGTTACTGGCTGGAGATGAGATCGCGCATCAATGACATTGGTTC	117		
QY	251	tgagcagcttcagaaaaggagcagttccctcgttgaaagatcctcaagaccctcaggc	310		
Dd	118	TCTCAGGTTTAGACAGGAACGATTTGCCGTGTGAAGGGTGCCATAAGACCCTGAAGC	177		
QY	311	tggtagtcgcgaaaaatggggctctgtctatgttgccccagaat	352		
Dd	178	TGGTCGTCAAAGGAGGAGCGAGCTGGGCTTGGAGGCCCTCACT	219		

RESULT 8

BF434691 558 bp mRNA EST 29-NOV-2000
LOCUS 7p03b06.x1 NCI_CGAP_Ov18 Homo sapiens cDNA clone IMAGE:3644507 3'
DEFINITION similar to SW:APXL_HUMAN Q13796 APICAL-LIKE PROTEIN ;, mRNA
sequence.

ACCESSION BF434691
VERSION BF434691.1 GI:11446979
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 558)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
JOURNAL Contact: Robert Strausberg, Ph.D.
COMMENT Email: cgapsb@mail.nih.gov
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
R. Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: M. Bento
Soares, Ph.D. cDNA Library Arrayed by: Christa Prange, The
I.M.A.G.E. Consortium DNA Sequencing by: Washington University
Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL, send email to:
info@image.llnl.gov
Seq primer: -40UP from Gibco
High quality sequence stamp 488.

FEATURES

SOURCE

Location/Qualifiers
1. .558

BASE COUNT
ORIGIN

```

/lab.host="DH10B (phage-resistant)"
/note="Organ: ovary; Vector: pT733-D-Pac (pharmacia) with a
modified polynukler; Site:1: Not I; Site2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTTACCATGTGATGATGGAGGAGCGCCGCGGCACATTTTTTTTTTTT 3']
double-stranded cDNA was ligated to Eco RI adaptors
(pharmacia), digested with Not I and cloned into the Not
I and Eco RI sites of the modified pT73 vector. library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo. "
191 c 147 g 102 t

```

Query Match	18.0%;	Score 72.2;	DB 11;	Length 558;
Best Local Similarity	62.0%;	Pred. No. 2.5e-09;		
Matches 132;	Conservative 0;	Mismatches 78;	Indels 3;	Gaps 1.

[illegible][illegible]

ONTOGENESIS OF THE
Eukaryota; Metazoa; Chordata; Craniata; Euteleostomi;
Mammalia; Eutheria; Primates; Carnivora; Hominoidea; Homo.
1 (pages 1 to 383)
Mallory, G., Wallace, J.C., Smith, K., Swartzell, S., Holman, T.,
Keller, A., Shaver, R., Furlong, J., Young, J., Zhao, S., Adams, M.D., and
REFERENCE
AUTHORS

TITLE	JOURNAL	MEDLINE	COMMENT
Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome	Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)	99380589	Contact: Mahairas GG, Wallace JC, Hood L High Throughput Sequencing Center

FEATURES

SOURCE

LOCATION/QUANTITY
1. .383

BASE COUNT	86 a	92 c	115 g	90 t	
ORIGIN	/note="Organ: sperm; Vector: pBeloBAC11; BAC clones in E-coli DH10B"				

Query Match	16.1%;	Score 64.6;	DB 13;	Length 383;
Best Local Similarity	65.7%;	Pred. No. 2.6e-07;		
Matches 94; Conservative	0;	Mismatches 49;	Indels 0;	Gaps

[illegible]

DB 276 ATCATGACATGGCTCTCAGGGTTTACAGGAAAGCATTTGCTGTGAGGGGGTCC 335
 QY 295 tacaagaccctcagctggtgtgt 317
 DB 336 CATAAGACCCCTGAAGCTGCTGT 358

RESULT	10						
AM001038/c							
LOCUS	AM001038	354 bp	mRNA				08-MAR-2000
DEFINITION	w91h03.x1 NCI_CGAP_kid1 Homo sapiens CDNA clone IMAGE:249509 similar to SW:AFXL_HUMAN Q13796 APICAL-LINE PROTEIN ; mRNA sequence.						
ACCESSION	AM001038						
VERSION	AM001038.1	GI:5847954					
KEYWORDS	EST.						
SOURCE	human.						
ORGANISM	Homo sapiens						

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 354)
NCI-CCGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP)
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strussberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael
Imbert-Rock, M.D., Ph.D.

FEATURES	location/Qualifiers
source	1..354
	/organism="Homo sapiens"
	/db_xref="taxon:9606"

```

/clone="IMAGE:2495093"
/clone_lib="NCI-CGAP_Kid1"
/lab_host="DH10B"
/note="Organ: Kidney; Vector: pTY73D-Pac (Pharmacia)
a modified polylinker; Site_1: Not I; Site_2: Eco RI;
plasmid DNA from the normalized library NCI-CGAP_Kid1
prepared, and ss circles were made in vitro. Following
purification, this DNA was used as tracer in a Southern
hybridisation reaction. The driver was PCR-amplified

```

BASE COUNT	96 a	48 c	70 g	94 t	1 others
ORIGIN					

Oy	322	agaaatggggtctgtcatgtttgccccagaaatggaagtgatgtgcatccatcatgatat	381
Db	232	AGAAATGGGCTCTTGGTATGTTCTTTACGGCTGGAATGCAAGTGTATTACAGGCATCAT	172
Oy	382	catcatgcaactgcagaccttga	402
Db	172	TATAGCGCACTACAGCCTTGA	152

RESULT	15
A1687027	
LOCUS	463 bp
DEFINITION	mRNA EST 14-DEC-1999
ACCESSION	t929db03.x1 NCI-CGAP-DT3 Homo sapiens cDNA clone IMAGE:2206733
VERSION	similar to contains Alu repetitive element;; mRNA sequence.
KEYWORDS	A1687027 A1687027 GI:4896321
	EST.

Fri Feb 1 08:56:34 2002

SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 463)
 AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs@mail.nih.gov
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
 Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: Life Technologies, Inc.
 DNA Sequencing by: Greg Lennon, Ph.D.
 CDNA Library Arrayed by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
 Insert Length: 1288 Std Error: 0.00
 Seq primer: -400P from Gibco
 High quality sequence stop: 418.
 FEATURES Location/Qualifiers
 1..463
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:2206733"
 /clone_lib="NCI-CGAP-Ut3"
 /tissue_type="poorly-differentiated endometrial
 adenocarcinoma, 2 pooled tumors"
 /lab_host="DH10B"
 /note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: SalI;
 Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
 Average insert size 1.45 kb. Life Technologies catalog #:
 11541-018"
 BASE COUNT 144 a 105 c 87 g 125 t 2 others
 ORIGIN
 Query Match 14.6%; Score 58.8; DB 10; Length 463;
 Best Local Similarity 84.6%; Pred. No. 1e-05;
 Matches 66; Conservative 0; Mismatches 12; Indels 0; Gaps 0;
 QY 322 agaaatgggggtttgtatgttgcacgaatgaaggtagtggtattcatgagcatgat 381
 ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 2 AGAGACAGGGTCTTGCTATGTGCCAGGCTGAGTGTAGTGGCTATTTCACAGCAGCAT 61
 QY 382 catcatgcactgcagcct 399
 ||| ||||| |||||
 Db 62 CATAGTGCACGCTT 79
 Search completed: January 31, 2002, 19:16:15
 Job time: 1963 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 31, 2002, 19:46:38 ; Search time 1599.5 Seconds
(without alignments)
4146.212 Million cell updates/sec

Title: US-09-641-831-5

Sequence: 402
1 atgatgagagaccactgaaga.....atcatgcactgcagccttga 402

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl: *
1: gb_ba: *
2: gb_hlg: *
3: gb_in: *
4: gb_om: *
5: gb_ov: *
6: gb_pat: *
7: gb_ph: *
8: gb_pl: *
9: gb_pr: *
10: gb_ro: *
11: gb_sts: *
12: gb_sy: *
13: gb_un: *
14: gb_vl: *
15: em_ba: *
16: em_fun: *
17: em_hum: *
18: em_in: *
19: em_om: *
20: em_ov: *
21: em_ov: *
22: em_pat: *
23: em_ph: *
24: em_pl: *
25: em_ro: *
26: em_sts: *
27: em_sy: *
28: em_un: *
29: em_vl: *
30: em_hlgo_hum: *
31: em_hlgo_inv: *
32: em_hlgo_rod: *
33: em_hlgo_hum: *
34: em_hlgo_inv: *
35: em_hlgo_rod: *
36: em_hlgo_other: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	402	100.0	402	6	AX088027 Sequence
2	323	80.3	327	6	AX088025 Sequence
3	323	80.3	573	6	AX088023 Sequence
4	211.2	52.5	6423	10	AF199421 Mus muscu
5	170	42.3	163535	2	AC046189 Homo sapi
6	158.6	39.5	156375	2	AP002859 Homo sapi
7	98.8	24.6	7445	9	HSAPXL
8	85.2	21.2	6014	9	AB033028 Homo sapi
9	81	20.1	156375	2	AP002859 Homo sapi
10	64.6	16.1	196542	9	AL512843 Homo sapi
11	63.6	15.8	75684	2	AL1390030 Homo sapi
12	62.8	15.6	87451	2	AL1390030 Homo sapi
13	62	15.4	127639	2	HS079963
14	62	15.4	168957	2	AC025079
15	61.8	15.3	135693	9	HSJ116744
16	61.6	15.2	160770	2	AC016796
17	61	15.2	112664	9	AC020719
18	61	15.2	112664	9	HS12803
19	61	15.2	112664	9	AC031989
20	60.6	15.1	193126	2	CNS018P3
21	60.2	15.0	45262	9	AL391689
22	60.2	15.0	45334	9	HS1057D18
23	60.2	15.0	123331	2	AC010942
24	60.2	15.0	106687	9	AC005164
25	60.2	15.0	140730	2	AC011886
26	60.2	15.0	157103	2	AC0024947
27	60.2	15.0	163031	2	AC004958
28	60.2	15.0	163889	2	AL451010
29	60.2	15.0	183480	2	AL390040
30	60.2	15.0	186350	2	AC074325
31	60.2	15.0	202304	2	AC079240
32	60.2	15.0	202324	2	AC016397
33	60.2	15.0	202324	2	AC016397
34	59.4	14.8	82601	9	AL136971
35	59.4	14.8	142441	2	HSJ344H20
36	59.4	14.8	170242	2	AC018545
37	59.2	14.7	190030	2	AC079121
38	59.2	14.7	195881	9	CNS00007
39	59.2	14.7	195881	9	AC007055
40	59	14.7	199927	9	AL365505
41	59	14.7	101500	9	AL365505
42	59	14.7	171812	9	AC013399
43	58.8	14.6	42213	9	AC012447
44	58.8	14.6	42213	9	AC010527
45	58.8	14.6	156294	2	AC092409
					AC067984 Homo sapi

ALIGNMENTS

RESULT 1
AX088027
LOCUS AX088027
DEFINITION Sequence 5 from Patent WO0114422.
ACCESSION AX088027
VERSION AX088027.1 GI:13396953
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
AUTHORS Turner,C.A., Zambrowicz,B., Friedrich,G., Nehls,M. and Sands,A.T.
TITLE Sequence derived from a human mammary gland cdna library
JOURNAL Patent: WO 0114422-A 5 01-MAR-2001;
Lexicon Genetics Incorporated (US)
FEATURES
source 1..402
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"


```
BASE COUNT      110 a      91 c      117 g      84 t
ORIGIN

Query Match      100.0%; Score 402; DB 6; Length 402;
Best Local Similarity 100.0%; Pred. No. 2.2e-110; Mismatches 0; Indels 0; Gaps 0;
Matches 402; Conservative 0;

Qy 1 atgatgaggaccactgaagacttcacaaagcctagtgccacattaaactctaaacacggcc 60
Db 1 ATGATGAGGACCCTGAAGACTTCCAAAGCCTAGTGCCACATTAAACTCTAACACGGCC 60

Qy 61 accaaggaaggtacattatctgagggcattcctgagggagagctccctgggtttt 120
Db 61 ACCAAGGAAGGTACATTATCTGAGGCATCTCTGAGGGAGGAGCTCCCTGGGGTTTT 120

Qy 121 actctaaagggtggcctggagcagcgagaaaccattaatcatctctaaaggaagg 180
Db 121 ACTCTAAAGGTTGGCCTGGAGCAGCGAGAACCACTTAATCATCTCTTAAGGTCGAAGAAGG 180

Qy 181 ggcataagcagacacccctgagctccaaactgagctgggatgaggttggcacaat 240
Db 181 GGCAAGCAGACACCCCTGAGCTCCAAACTGCAGCTGGGATGAGGTGTGCACATCAAT 240

Qy 241 gaggtagcttgagcagctccagaaagggagcagtttccctgggtgaaaggattcctacaag 300
Db 241 GAGGTGACTCTGAGCAGCTCCAGAAAGGAGCGAGTTTCCCTGGTGAAGGATCCTACAAG 300

Qy 301 accctcaggtggtagtcgcag 323
Db 301 ACCCTCAGGCTGGTAGTGGCGAG 323

RESULT 3
AX088023 573 bp DNA PAT 17-MAR-2001
LOCUS
DEFINITION Sequence 1 from Patent WO0114422.
ACCESSION AX088023
VERSION AX088023.1 GI:13396951
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 573)
AUTHORS Turner,C.A., Zambrowicz,B., Friedrich,G., Nehls,M. and Sands,A.T.
TITLE Sequence derived from a human mammary gland cdna library
JOURNAL Patent: WO 0114422-A 1 01-MAR-2001;
Lexicon Genetics Incorporated (US)
FEATURES
Source 1..573
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 151 a 152 c 152 g 118 t
ORIGIN

Query Match 80.3%; Score 323; DB 6; Length 573;
Best Local Similarity 100.0%; Pred. No. 1.4e-86; Mismatches 0; Indels 0; Gaps 0;
Matches 323; Conservative 0;

Qy 1 atgatgaggaccactgaagacttcacaaagcctagtgccacattaaactctaaacacggcc 60
Db 1 ATGATGAGGACCCTGAAGACTTCCAAAGCCTAGTGCCACATTAAACTCTAACACGGCC 60

Qy 61 accaaggaaggtacattatctgagggcattcctgagggagagctccctgggtttt 120
Db 61 ACCAAGGAAGGTACATTATCTGAGGCATCTCTGAGGGAGGAGCTCCCTGGGGTTTT 120

Qy 121 actctaaagggtggcctggagcagcgagaaaccattaatcatctctaaaggaagg 180
Db 121 ACTCTAAAGGTTGGCCTGGAGCAGCGAGAACCACTTAATCATCTCTTAAGGTCGAAGAAGG 180

Qy 181 ggcataagcagacacccctgagctccaaactgagctgggatgaggttggcacaat 240
Db 181 GGCAAGCAGACACCCCTGAGCTCCAAAGCCTAGTGCCACATTAAACTCTAACACGGCC 240

Qy 241 gaggtagcttgagcagctccagaaagggagcagtttccctgggtgaaaggattcctacaag 300
Db 241 GAGGTGACTCTGAGCAGCTCCAGAAAGGAGCGAGTTTCCCTGGTGAAGGATCCTACAAG 300

Qy 301 accctcaggtggtagtcgcag 323
Db 301 ACCCTCAGGCTGGTAGTGGCGAG 323

RESULT 4
AF199421 6423 bp mRNA ROD 13-DEC-1999
LOCUS
```

```
BASE COUNT      110 a      91 c      117 g      84 t
ORIGIN

Query Match      100.0%; Score 402; DB 6; Length 402;
Best Local Similarity 100.0%; Pred. No. 2.2e-110; Mismatches 0; Indels 0; Gaps 0;
Matches 402; Conservative 0;

Qy 1 atgatgaggaccactgaagacttcacaaagcctagtgccacattaaactctaaacacggcc 60
Db 1 ATGATGAGGACCCTGAAGACTTCCAAAGCCTAGTGCCACATTAAACTCTAACACGGCC 60

Qy 61 accaaggaaggtacattatctgagggcattcctgagggagagctccctgggtttt 120
Db 61 ACCAAGGAAGGTACATTATCTGAGGCATCTCTGAGGGAGGAGCTCCCTGGGGTTTT 120

Qy 121 actctaaagggtggcctggagcagcgagaaaccattaatcatctctaaaggaagg 180
Db 121 ACTCTAAAGGTTGGCCTGGAGCAGCGAGAACCACTTAATCATCTCTTAAGGTCGAAGAAGG 180

Qy 181 ggcataagcagacacccctgagctccaaactgagctgggatgaggttggcacaat 240
Db 181 GGCAAGCAGACACCCCTGAGCTCCAAACTGCAGCTGGGATGAGGTGTGCACATCAAT 240

Qy 241 gaggtagcttgagcagctccagaaagggagcagtttccctgggtgaaaggattcctacaag 300
Db 241 GAGGTGACTCTGAGCAGCTCCAGAAAGGAGCGAGTTTCCCTGGTGAAGGATCCTACAAG 300

Qy 301 accctcaggtggtagtcgcag 323
Db 301 ACCCTCAGGCTGGTAGTGGCGAG 323

RESULT 3
AX088023 573 bp DNA PAT 17-MAR-2001
LOCUS
DEFINITION Sequence 1 from Patent WO0114422.
ACCESSION AX088023
VERSION AX088023.1 GI:13396951
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 573)
AUTHORS Turner,C.A., Zambrowicz,B., Friedrich,G., Nehls,M. and Sands,A.T.
TITLE Sequence derived from a human mammary gland cdna library
JOURNAL Patent: WO 0114422-A 1 01-MAR-2001;
Lexicon Genetics Incorporated (US)
FEATURES
Source 1..573
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 151 a 152 c 152 g 118 t
ORIGIN

Query Match 80.3%; Score 323; DB 6; Length 573;
Best Local Similarity 100.0%; Pred. No. 1.4e-86; Mismatches 0; Indels 0; Gaps 0;
Matches 323; Conservative 0;

Qy 1 atgatgaggaccactgaagacttcacaaagcctagtgccacattaaactctaaacacggcc 60
Db 1 ATGATGAGGACCCTGAAGACTTCCAAAGCCTAGTGCCACATTAAACTCTAACACGGCC 60

Qy 61 accaaggaaggtacattatctgagggcattcctgagggagagctccctgggtttt 120
Db 61 ACCAAGGAAGGTACATTATCTGAGGCATCTCTGAGGGAGGAGCTCCCTGGGGTTTT 120

Qy 121 actctaaagggtggcctggagcagcgagaaaccattaatcatctctaaaggaagg 180
Db 121 ACTCTAAAGGTTGGCCTGGAGCAGCGAGAACCACTTAATCATCTCTTAAGGTCGAAGAAGG 180

Qy 181 ggcataagcagacacccctgagctccaaactgagctgggatgaggttggcacaat 240
Db 181 GGCAAGCAGACACCCCTGAGCTCCAAAGCCTAGTGCCACATTAAACTCTAACACGGCC 240

Qy 241 gaggtagcttgagcagctccagaaagggagcagtttccctgggtgaaaggattcctacaag 300
Db 241 GAGGTGACTCTGAGCAGCTCCAGAAAGGAGCGAGTTTCCCTGGTGAAGGATCCTACAAG 300

Qy 301 accctcaggtggtagtcgcag 323
Db 301 ACCCTCAGGCTGGTAGTGGCGAG 323

RESULT 4
AF199421 6423 bp mRNA ROD 13-DEC-1999
LOCUS
```

QY	3	gatgagagacacactggaagactccacaagcctagtgcacattaatctcaacgycac	62
Db	232	GATGAAGACTCCGGAGAACTTGGAGAGACCCTACGACCCAAACCCAGTAGGACCC	291
QY	63	caaggaaggtacattctctggaggcaattctcggaggaggaagcctccctgggttttc	122
Db	292	CAGGGAGAAATTCGTTTATCTGGAAAGCCCTCTTAGAGAGAGGGCTCCCTGGGGCTTCAC	351
QY	123	tcctaaaggtgcctcggagacacgagaaaccattaatcctctaagtcgaaagagggg	182
Db	352	CTGGAAGAGGTGGCTGTGAGACCGTGAGAACCCGTTATCATCTTCCAAAGATTAAGAAAGGGG	411
QY	183	caagcagacacccctgagcttccaaactcagagctcggggatgaagttgtgcacaatga	242
Db	412	CAAAAGCCGACTAGTAGCTCCGGAGCTGAAAGCTGGGAGCAAAATCATACATCATATGA	471
QY	243	ggtgactcggagcagcactcccaaaaagagagcagtttccctggtgaagaatcctacaagc	302
Db	472	GGTGGGGCTGAGACAGTCCAGAGAGAGAGGCTGTCTCCCTAGTGAAGGGCTCATAGAC	531
QY	303	cctcaggtcgtgtagtgcgcagaatggg	330
Db	532	CTGCGCGCTGTGTGTGCGCAGAGACGTG	559
RESULT	5		
AC064189	163535 bp	DNA	23-SEP-2000
LOCUS			
DEFINITION	Homo sapiens chromosome 4 clone RP11-356M17 map 4, WORKING DRAFT		
ACCESSION	AC064189		
VERSION	AC064189.4		
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	1 (bases 1 to 163535)		
JOURNAL	Birten, B., Linton, L., Nusbaum, C. and Lander, E.		
TITLE	Homo sapiens chromosome 4, clone RP11-356M17		
REFERENCE	Unpublished		
AUTHORS	2 (bases 1 to 163535)		
JOURNAL	Birten, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N., Anderson, S., Baldwin, J., Bana, N., Bastien, V., Beda, F., Boguslavsky, L., Boukhalter, B., Brown, A., Burkett, G., Campiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P., Deatellano, K., Dewar, K., Diaz, J. S., Dodge, S., Domino, K., Doyle, M., Ferrer, M., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L., Grand-Pierre, N., Grant, G., Hagos, B., Heald, A., Horton, L., Howland, J. C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karitas, A., Klein, J., Larocque, K., Lamazars, R., Landers, T., Lechoczy, U., Levine, R., Lieu, C., Liu, G., Locke, K., MacDonald, P., McPheters, N., McCarthy, M., Medwan, P., McGura, A., McKernan, K., Morrow, J., Murphy, T., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neil, D., Olyar, T. M., Oliver, J., Peterson, K., Pierre, N., Pisan, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B., Stange-Rhomon, N., Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Triggillo, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, V., Zimmer, A. and zody, M.		
TITLE	Direct Submission		
JOURNAL	Submitted (13-APR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA		
COMMENT	On Sep 10, 2000 this sequence version replaced g1:8083636. All repeats were identified using RepeatMasker: Smit, A.P.A. & Green, P. (1996-1997) http://ftp.genome.washington.edu/RM/RepeatMasker.html		
	Genome Center		

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIGR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence.submissions@genome.wi.mit.edu

----- Project Information

Center project name: L9512

Center clone name: 356_M.17

----- Summary Statistics

Sequencing vector: M13; M77815; 100% of reads

Chemistry: Dye-terminator; Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 148971 bases at least Q40

Consensus quality: 156340 bases at least Q30

Consensus quality: 159016 bases at least Q20

Insert size: 181000; agarose-fp

Insert size: 160935; sum-of-contigs

Quality coverage: 3.6 in Q20 bases; agarose-fp

Quality coverage: 4.0 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently consists of 27 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

```

* 1 2095: contig of 2095 bp in length
* 2096 2195: gap of 100 bp
* 2196 3520: contig of 1325 bp in length
* 3521 3620: gap of 100 bp
* 3621 4777: contig of 1157 bp in length
* 4778 4877: gap of 100 bp
* 4878 6660: contig of 1783 bp in length
* 6661 6760: gap of 100 bp
* 6761 8530: contig of 1770 bp in length
* 8531 8630: gap of 100 bp
* 8631 11149: contig of 2519 bp in length
* 11150 11249: gap of 100 bp
* 11250 15298: contig of 4049 bp in length
* 15299 15398: gap of 100 bp
* 15399 18533: contig of 3135 bp in length
* 18534 18633: gap of 100 bp
* 18634 22044: contig of 3411 bp in length
* 22045 22144: gap of 100 bp
* 22145 25578: contig of 3434 bp in length
* 25579 25678: gap of 100 bp
* 25679 29571: contig of 3893 bp in length
* 29572 29671: gap of 100 bp
* 29672 34405: contig of 4734 bp in length
* 34406 34505: gap of 100 bp
* 34506 37415: contig of 2910 bp in length
* 37416 37515: gap of 100 bp
* 37516 44162: contig of 6647 bp in length
* 44163 44262: gap of 100 bp
* 44263 48837: contig of 4575 bp in length
* 48838 48937: gap of 100 bp
* 48938 55115: contig of 6178 bp in length
* 55116 55215: gap of 100 bp
* 55216 62194: contig of 6979 bp in length
* 62195 62294: gap of 100 bp
* 62295 69547: contig of 7253 bp in length
* 69548 69647: gap of 100 bp
* 69648 78050: contig of 8403 bp in length
* 78051 78150: gap of 100 bp
* 78151 88041: contig of 9891 bp in length
* 88042 88141: gap of 100 bp
* 88142 98560: contig of 10419 bp in length
* 98561 98660: gap of 100 bp
* 98661 108146: contig of 9486 bp in length
* 108147 108246: gap of 100 bp
* 108247 118855: contig of 10609 bp in length
* 118856 118955: gap of 100 bp

```

```

* 118956 128543: contig of 9588 bp in length
* 128544 128643: gap of 100 bp
* 128644 142647: contig of 14004 bp in length
* 142648 142747: gap of 100 bp
* 142748 153569: contig of 10822 bp in length
* 153570 153669: gap of 100 bp
* 153670 163535: contig of 9866 bp in length.

```

Location/Qualifiers

1. 163535

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="4"

/map="4"

/clone_lib="RPC1-11 Human Male BAC"

1. 2095

/note="assembly_fragment"

clone_end:SP6

vector_side:left"

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

BASE COUNT 48438 a 32205 c 31514 g 48770 t 2608 others
ORIGIN

Query Match 42.3%; Score 170; DB 2; Length 163535;
Best Local Similarity 100.0%; Pred. No. 2.5e-40;
Matches 170; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 atgataggagcactgaagcattccacaagcctgtccacattaaacttaacagggc 60
62580 ATGATGAGGACACTGAAGACTTCACAGGCTAGTCCACATTAACCTTAACACGGCC 62539

Db 61 accaagggaaggtacattatcttgaggcattcctgtgaggaggagagctccctggggttt 120
62640 ACCAAGGGAAGGTACTTATCTGAGGACTTCTCTGAGGAGAGAGCTCCCTGGGGTTT 62639

Db 121 accttaagggtgctctgagacagcaggaaccatcatcatcctaagt 170
62700 ACTCTAAGGTTGCTCTGAGACAGGAGACCATTAATCATCTTAAGT 62749

RESULT 6
AP002859 156375 bp DNA HTG 29-AUG-2000
LOCUS Homo sapiens chromosome 4 clone 225E15 map 4q16-q20, *** SEQUENCING
DEFINITION IN PROGRESS ***, 15 unordered pieces.
ACCESSION AP002859
VERSION AP002859.1 GI:9955382
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Homo sapiens DNA, clone:225E15.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 156375)
Tsal,S.F.
Direct Submission
Submitted (28-AUG-2000) Shih-Feng Tsal, National Yang-Ming
University, Institute of Genetics, 155 Li-Kong St. Section 2,
Petou, Taipei, Taiwan 11221, Republic of China
(E-mail:ymptsa@ym.edu.tw, URL:http://genome.ym.edu.tw/
Tel:886-2-28267043, Fax:886-2-28264930)
These sequences are draft human sequences, not finished sequences.
These sequences are represented as 100 N.
Of the same clone are represented as 100 N.
NOTE: This is a 'working draft' sequence. It currently
consists of 15 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.

1 21185: contig of 21185 bp in length
21186 21285: gap of 100 bp
21286 22928: contig of 1643 bp in length
22929 23028: gap of 100 bp
23029 26285: contig of 3257 bp in length
26286 26385: gap of 100 bp
26386 30606: contig of 4221 bp in length
30607 30706: gap of 100 bp
30707 36924: contig of 6218 bp in length
36925 37024: gap of 100 bp
37025 40953: contig of 3929 bp in length
40954 41053: gap of 100 bp
41054 47843: contig of 6790 bp in length
47844 47943: gap of 100 bp
47944 57174: contig of 9231 bp in length
57175 57274: gap of 100 bp
57275 64994: contig of 7720 bp in length
64995 65094: gap of 100 bp
65095 76523: contig of 11429 bp in length
76524 76623: gap of 100 bp
76624 88306: contig of 11683 bp in length

FEATURES
source
1. 156375
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="4"
/map="4q16-q20"
/clone="225E15"

BASE COUNT 45683 a 32392 c 31715 g 45174 t 1411 others
ORIGIN

Query Match 39.5%; Score 158.6; DB 2; Length 156375;
Best Local Similarity 97.6%; Pred. No. 6.9e-37;
Matches 161; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Db 167 agtgcgaagagggaagcagacacccctgagctcccaactgcagctgggagtagg 226
80640 AGTGCAGAGGAGGAGGAGCAGACACCTGAGCTCCAACTGCAGGCTGGGATGAGG 80699

Db 227 ttgtgcacatcaatgagtgctctgagcagctccagaagaggagcttccctgtga 286
80700 TTGTGCACATCAATGAGTGTGCTGAGCAGCTCCAGAAAGAGAGAGTTCCCTGTGTA 80759

Db 287 aagatcctacaagaccctcagctgtagtgcgcaagaatgggg 331
80760 AAGATCTCTACAGACCTCAGGCTGTAGTGCAGAGTAGGTGG 80804

RESULT 7
HSAPXL 7445 bp mRNA PRI 10-DEC-1995
LOCUS H. sapiens APXL mRNA.
DEFINITION X83543
ACCESSION X83543.1 GI:790999
VERSION X83543.1 GI:790999
KEYWORDS APXL gene.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 7445)
Schiaffino,M.V., Bassi,M.T., Rugarli,E.I., Renieri,A., Galli,L. and
Ballabio,A.
Cloning of a human homologue of the Xenopus laevis Arx gene from
the ocular albinism type 1 critical region
Hum. Mol. Genet. 4 (3), 373-382 (1995)
JOURNAL 95315933
MEDLINE 2 (bases 1 to 7445)
REFERENCE Schiaffino,V.M.
AUTHORS Direct Submission
JOURNAL Submitted (19-DEC-1994) V.M. Schiaffino, T.I.G.E.M., via Olgettina
58, 20132 Milano, ITALY
FEATURES
source
1. 7445
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="22"
/dev_stage="adult"
/tissue_type="retina"
/map="xp22.2-22.3"
91.4941
/gene="APXL"
91.4941
/gene="APXL"
/codon_start=1
/protein_id="CAA58534.1"

NOTE: This is a 'working draft' sequence. It currently consists of 15 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1	21185:	contig of 21185 bp in length
2	21186	21285: gap of 100 bp
3	21286	22928: contig of 1643 bp in length
4	22929	23028: gap of 100 bp
5	23029	26285: contig of 3357 bp in length
6	26286	26385: gap of 100 bp
7	26386	30606: contig of 4221 bp in length
8	30607	30706: gap of 100 bp
9	30707	36924: contig of 6218 bp in length
10	36925	37024: gap of 100 bp
11	37025	40953: contig of 3929 bp in length
12	40954	41053: gap of 100 bp
13	41054	47843: contig of 6790 bp in length
14	47844	47943: gap of 100 bp
15	47944	57174: contig of 9231 bp in length
16	57175	57274: gap of 100 bp
17	57275	64994: contig of 7720 bp in length
18	64995	65094: gap of 100 bp
19	65095	76523: contig of 11429 bp in length
20	76524	76623: gap of 100 bp
21	76624	88306: contig of 11683 bp in length
22	88307	88406: gap of 100 bp
23	88407	101162: contig of 12756 bp in length
24	101163	101262: gap of 100 bp
25	101263	145436: contig of 14174 bp in length
26	115437	115536: gap of 100 bp
27	115537	133350: contig of 17814 bp in length
28	133351	133450: gap of 100 bp

TITLE	Direct Submission
JOURNAL	Submitted (20-AUG-1997) Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE	4 (bases 1 to 196542)
AUTHORS	Chiu, M.W.
TITLE	Direct Submission
JOURNAL	Submitted (04-SEP-1997) Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, Texas 77030, USA
REFERENCE	5 (bases 1 to 196542)
AUTHORS	Worley, K. C.
TITLE	Direct Submission
JOURNAL	Submitted (22-AUG-1998) Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Sep 4, 1997 this sequence version replaced gi:1766071 gi:1773380
 gi:2289024.
 Sequencing is completed to a minimum standard of double strand
 coverage with a minimum of 2 clones and 2 reads with no ambiguities
 or 2 chemistries with a minimum of 2 clones and 3 reads with no
 ambiguities. If the sequence quality does not meet this standard,
 it will be indicated in the annotation.

The repeat regions shown were identified using RepeatMasker by
 Adrian Smit.
 Sequence similarities were identified using Powerblast by Jinghui
 Zhang.

Exon/Intron boundaries of identified genes were chosen if there
 were canonical splice junctions that maintained sequence continuity
 across the splice junctions:

FEATURES
 source Location/Qualifiers
 1. .196542
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="U177G4"
 /chromosome="X"
 /clone.lib="Livermore Chromosome X Cosmid Library
 LLOXNC01"
 /map="Xp22"
 complement(760..1057)
 /rpt_family="AluJb"
 1893..2194
 /rpt_family="AluSg"
 2305..2584
 /rpt_family="AluSx"
 3418..3630
 /rpt_family="MER20"
 complement(3806..4096)
 /rpt_family="AluSx"
 4138..4537
 /rpt_family="L1MB2"
 complement(4549..4841)
 /rpt_family="AluJo"
 complement(5421..5528)
 /rpt_family="MER58B"
 complement(7001..7132)
 /rpt_family="FAM"
 7773..7903
 /rpt_family="AluJ"
 7910..8215
 /rpt_family="AluYa5"
 8224..8401
 /rpt_family="AluJ"
 complement(8932..9232)
 /rpt_family="AluY"
 complement(9458..9481)
 /rpt_family="POLY_A"
 11579..11618
 /rpt_family="AT-rich"
 complement(11619..11916)
 /rpt_family="AluSg"
 complement(11917..12252)
 /rpt_family="AluJb"
 complement(12328..12633)
 /rpt_family="AluSx"
 complement(12665..13117)
 /rpt_family="L1MB3"
 complement(13236..13267)
 /rpt_family="L1MA1"
 13323..13604
 /rpt_family="AluJb"
 13983..14378
 /note="Similar to sequence with GenBank Accession number
 236840"

repeat_region
 /rpt_family="MER58B"
 15565..15565
 repeat_region
 /rpt_family="AluSg"
 15566..15566
 repeat_region
 /rpt_family="L1MB7"
 16091..16660
 mRNA
 join(<17023..17176,34350..34481,37731..40071,41563..41663,
 75548..76243,80507..81058,82568..82739,88014..88286,
 90044..>92814)
 /gene="HSAPXL"
 /product="APXL"
 join(<17023..17176,34350..34481,37731..40071,41563..41663,
 75548..76243,80507..81058,82568..82739,88014..88286,
 90044..>92814)
 /gene="HSAPXL"
 gene
 <17023..>92814
 /gene="HSAPXL"
 gene
 join(<17023..17176,34350..34481,37731..40071,41563..41663,
 75548..76243,80507..81058,82568..82739,88014..88286,
 90044..90310)
 /gene="HSAPXL"
 to sequence encoded by GenBank Accession Number X83543"
 /codon_start=3
 /product="APXL"
 /protein_id="AAC32592.1"
 /db_xref="GI:1773381"
 /translation="IEGSKAAAVDKLIAGDEIVGINDIGLSGFRQEAICLVKSGHKT
 LKLVKRSSELGWRPHSKFSDSHPELAASFTSTGCPSSWGRHHSSSSHDLS
 SSWEQTNLQTLDFHSSGLSDHPSRLSVAKSNSSIDHLSGSKRDSAYGSFST
 SSSTPDHLSKADTSSANILYTVGLWEAPRQGGQAAGDPGSGEKLSCFPVRPV
 HGSGKPRPEYNAEPKLAAPGRSNFPGVYVFDKRAKPSPPPPPLRSDFPAATKS
 HKAQPVFEAAAQHTALAAQAPGRDPRPDLTDRPWRSAHSGKSGGGGCGCPQ
 EAHADGSPKDKGASRLQSLSSDYRFPQSPHGRHPLYSDHSLPCADSLGQEP
 GAASFONDSPPOVRLSSCDOKLGGWQPCVQGDQAQAQWAGWPSDTALGAE
 SLPTVGSQSPHLPQEPDARETGRCTVLDKAGECGAGAQEPAPRAAKASQ
 RLAAITWADGESSRICQETPLHSLTQEGKRRPESPESATRPFPDAHVGKPTR
 RSDRATTLRNEIQHRAKQSRSTVALTAGEAEDGTGRAGLGGTQEGPLAGT
 YKHLKEAQAQVLRATSEKRLDLPNPDGSELEHRMGDPDVPVPHWEAGLAQPPS
 STSGGPHPRIGRRRTAEQKLSYSEPKMNEVGLTRGYSPOHPRSTSEDTGTF
 DRWKEETKVPQPAQKALHGIPDKRPPRTAGTCEGTEPWRSTSLGDSLN
 AHSAAKAGTSDLPRLGTFAEYOASKEQKPLEARSGRCHSADDILVSLDQER
 POHVGRSSSTDYHQAESVELRQAGDPGEPEELPSAVRAEGQSTPQADQA
 CREGSPSQHPPSQKAPNPTFSELSCRCGAPELPRGRCGACPLPDYRVSESTP
 ADLGPASGSPHARQDQSWPVSSALLSRSPVSSAQPDPTKATVCGRSGHVSQDA
 PVGLGRPTTPSPASLDVYVARSLSHSPVSFSAQPDPTKATVCGRSGHVSQDA
 SRPLPALLPQKHLRQATMERSRSPQAFAPKLTDPKPLLIQDSDTRIERVM
 DNNTVMVPIKIVHSESQEPEKESQSLACPAEPALPHGKDKQIKTISTEQFYSR
 FCLYFQGAEPAPRAQAPQPLGTQVPEKDRCTGPPGLSYMKAKETVEDLKSE
 ELAREIVGKDSLADILDSVAKITMDLMGIFPKDEHLEEAQORRKLKLPKIPSPR
 STEERKEEPSVFAAATNATNTYISTAPKABELLIKMDLQEQHEEDSGSLDHDIL
 SYKKELIESIKRQLQVREARESLLEDVQANTVGAEEVIAVKGVCPSFEDKRFMF
 IGDLDKVNLLSLSGRLARVENLNLDGASPGDRSLLEKORVLIQOQHEDAKELK
 ENLDRRIVFDILANYLSEESLADYHFVRKMSALIIEQRELEDKIHLGSQLKLL
 DSIQPERGK"
 complement(17976..18074)
 /rpt_family="H13"
 18162..18473
 /rpt_family="AluJb"
 19636..19717
 /rpt_family="MER21A"
 complement(19798..19937)
 /rpt_family="L1MB3"
 complement(20181..20327)
 /rpt_family="FAM"
 complement(22367..22401)
 /rpt_family="AT-rich"
 22406..22461
 /rpt_family="(GGAA)n"
 complement(22462..22627)
 /rpt_family="AluJo"
 complement(22638..22933)
 /rpt_family="AluSx"
 repeat_region
 complement(17976..18074)
 repeat_region
 18162..18473
 repeat_region
 19636..19717
 repeat_region
 complement(19798..19937)
 repeat_region
 complement(20181..20327)
 repeat_region
 complement(22367..22401)
 repeat_region
 22406..22461
 repeat_region
 complement(22462..22627)
 repeat_region
 complement(22638..22933)
 repeat_region
 complement(22638..22933)

repeat_region	/note="L2 repeat: matches 1228. .2345 of consensus" 38554. .38703
repeat_region	/note="L2 repeat: matches 2074. .2225 of consensus" 41312. .41373
repeat_region	/note="MIR repeat: matches 104. .176 of consensus" 41372. .42513
repeat_region	/note="L2 repeat: matches 1610. .2745 of consensus" 42530. .42693
repeat_region	/note="MIR repeat: matches 60. .247 of consensus" 44660. .44787
repeat_region	/note="MIR repeat: matches 20. .160 of consensus" 45322. .45573
repeat_region	/note="MER46C repeat: matches 46. .328 of consensus" 46112. .46244
repeat_region	/note="L2 repeat: matches 2576. .2709 of consensus" 46264. .46406
repeat_region	/note="L2 repeat: matches 2411. .2558 of consensus" 46520. .46943
repeat_region	/note="L2 repeat: matches 1702. .2148 of consensus" 46954. .47188
repeat_region	/note="L2 repeat: matches 1951. .2215 of consensus" 47240. .47315
repeat_region	/note="L2 repeat: matches 1789. .1857 of consensus" 47316. .47619
repeat_region	/note="AluSg repeat: matches 1. .306 of consensus" 47620. .47908
repeat_region	/note="L2 repeat: matches 1461. .1789 of consensus" 47938. .49004
repeat_region	/note="MER11C repeat: matches 1. .1057 of consensus" 49018. .49086
repeat_region	/note="L2 repeat: matches 1389. .1462 of consensus" 49660. .49750
repeat_region	/note="MLT1J repeat: matches 103. .199 of consensus" 51571. .51786
repeat_region	/note="MIR repeat: matches 35. .247 of consensus" 51747. .51971
repeat_region	/note="MIR repeat: matches 7. .224 of consensus" 52048. .52353
repeat_region	/note="AluX repeat: matches 1. .307 of consensus" 52905. .53006
repeat_region	/note="51 copies 2 mer to 60% conserved" 53361. .53422
repeat_region	/note="31 copies 2 mer ca 71% conserved" 53471. .53579
repeat_region	/note="L2 repeat: matches 2628. .2736 of consensus" 53754. .53916
repeat_region	/note="L1MC/D repeat: matches 5375. .5396 of consensus" 53917. .54205
repeat_region	/note="AluSg repeat: matches 1. .289 of consensus" 54206. .54500
repeat_region	/note="L1MC/D repeat: matches 5094. .5375 of consensus" 55056. .55280
repeat_region	/note="MIR repeat: matches 3. .262 of consensus" 55391. .55644
repeat_region	/note="MIR repeat: matches 26. .261 of consensus" 56355. .56637
repeat_region	/note="AluX repeat: matches 21. .311 of consensus" 56650. .56786
repeat_region	/note="L1PAl3 repeat: matches 5998. .6144 of consensus" 57041. .57355
repeat_region	/note="AluY8 repeat: matches 1. .315 of consensus" 57476. .57556
repeat_region	/note="L2 repeat: matches 778. .857 of consensus" 60400. .60703
repeat_region	/note="L2 repeat: matches 2404. .2750 of consensus" 60950. .61243
repeat_region	/note="AluYb repeat: matches 1. .294 of consensus" 61616. .61838
repeat_region	/note="MIR repeat: matches 5. .241 of consensus" 62351. .62581
repeat_region	/note="MIR repeat: matches 22. .262 of consensus" 62881. .63097
repeat_region	/note="MIR repeat: matches 18. .254 of consensus"

```

repeat_region 63440..63598
/note="MIR repeat: matches 102..252 of consensus"
repeat_region 63599..63840
/note="AluSq repeat: matches 67..307 of consensus"
repeat_region 64159..64563
/note="L2 repeat: matches 2273..2710 of consensus"
repeat_region 64868..64960
/note="MIR repeat: matches 90..200 of consensus"
repeat_region 65432..66100

Query Match 15.8%; Score 63.6; DB 9; Length 75684;
Best Local Similarity 76.5%; Pred. No. 2.8e-08;
Matches 78; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

Qy 301 accctcaggctgtagcgccagaaatggggtctgtctatgttgcgcagaatggaagcta 360
Db 31278 AGCCACCACCTATTATTATTAAAGAGATGGGGCTCTGCTATGTTGCCACGACTGAAATGCA 31337

Qy 361 gtggctattcaggctgatgcataatcgcactgcagccttga 402
Db 31338 GTGGCTATTACAGGCGCTGATGACTCACTGCAGCTTTGA 31379

RESULT 12
AL390030 87451 bp DNA HTG 20-JAN-2001
LOCUS Homo sapiens chromosome 1 clone RP4-700A9, *** SEQUENCING IN
DEFINITION PROGRESS ***, 9 unordered pieces.
ACCESSION AL390030
VERSION AL390030.2 GI:9368769
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 87451)
AUTHORS Pavitt,R.
TITLE Direct Submission
JOURNAL Submitted (20-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Jul 22, 2000 this sequence version replaced gi:9368315.
COMMENT ----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: dj700A9
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator ET-amersham; 10% of reads Chemistry:
Dye-terminator Big Dye; 89% of reads
Consensus quality: 82540 bases at least Q40
Consensus quality: 84030 bases at least Q30
Consensus quality: 85000 bases at least Q20
Insert size: 86651; sum-of-contigs
Insert size: 101143; 4.6% error; agarose-ff
Quality coverage: 4.16x in Q20 bases; sum-of-contigs Quality
coverage: 3.57x in Q20 bases; agarose-ff
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 9 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 26100: contig of 26100 bp in length
* 26101 26200: gap of 100 bp
* 26201 33791: contig of 7591 bp in length

```

```

* 33792 33891: gap of 100 bp
* 33892 43982: contig of 10091 bp in length
* 43983 44082: gap of 100 bp
* 44083 55959: contig of 11877 bp in length
* 55960 56059: gap of 100 bp
* 56060 65631: contig of 9572 bp in length
* 65632 65731: gap of 100 bp
* 65732 68376: contig of 2645 bp in length
* 68377 68476: gap of 100 bp
* 68477 72701: contig of 4225 bp in length
* 72702 72801: gap of 100 bp
* 72802 78093: contig of 5292 bp in length
* 78094 78193: gap of 100 bp
* 78194 87451: contig of 9258 bp in length.
FEATURES
Source
1. 87451
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP4-700A9"
/clone_1bp="RPCI-4"
1. 26100
/note="assembly-fragment:00202"
misc_feature
/clone_end:SP6
vector_side:left"
26201..33791
/note="assembly-fragment:00014"
fragment_chain:1"
33892..43982
/note="assembly-fragment:00047"
fragment_chain:1"
44083..55959
/note="assembly-fragment:00350"
fragment_chain:1"
56060..65631
/note="assembly-fragment:00701"
fragment_chain:1"
65732..68376
/note="assembly-fragment:00007"
68477..72701
/note="assembly-fragment:00092"
72802..78093
/note="assembly-fragment:00258"
fragment_chain:2"
78194..87451
/note="assembly-fragment:00602"
fragment_chain:2"
clone_end:"7"
vector_side:right"
BASE COUNT 26181 a 19311 c 17915 g 23233 t 811 others
ORIGIN
Query Match 15.6%: Score 62.8; DB 2; Length 87451;
Best Local Similarity 85.4%: Pred. No. 4.9e-08;
Matches 70; Conservative 0; Mismatches 12; Indels 0; Gaps 0;
QY 321 cagaatggggctctgtctatgtccagaatggaagtgatggtctatcatcaaggcatga 380
||||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
Db 57022 CAGACATGGGATCTCCCTGTGCTAGACTGACGACGATGGCTATTCACAGCGCAGA 57081
||||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
QY 381 tcatactgacatgcagccttga 402
||||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
Db 57082 TCATCATGCACCTACGACCTTGA 57103
||||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
RESULT 13
HSDJ799G3 127639 bp DNA PRI 04-APR-2001
LOCUS Human DNA sequence from clone RP4-799G3 on chromosome 1q42.11-42.3
DEFINITION Contains STS and GSSs, complete sequence.
ACCESSION AL078624
VERSION AL078624.24 GI:6272171

```

```

KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 127639)
REFERENCE
AUTHORS Howden,P.
TITLE Direct Submission
JOURNAL Submitted (23-DEC-1999) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Nov 5, 1999 this sequence version replaced gi:6249407.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C.elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chrl
RP4-799G3 is from the library RPCI-4 constructed at the Roswell
Park Cancer Institute by the group of Pieter de Jong. For further
details see http://bacpac.med.buffalo.edu/
VECTOR: pCYPAC2
This sequence is the entire insert of clone RP4-799G3 The true
right end of clone RP4-646B12 is at 100 in this sequence.
FEATURES
Source
Location/Qualifiers
1..127639
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/map="q42.11-42.3"
/clone="RP4-799G3"
/clone_1bp="RPCI-4"
3330..3381
/note="26 copies 2 mer ag 82 conserved"
3708..4111
/note="match: GSS: Em:AQ733797"
4113..4243
/note="Alusg/x repeat: matches 2..133 of consensus"
6668..7101
/note="Tiger2a repeat: matches 1..434 of consensus"
7205..7481
/note="Ltr16B repeat: matches 107..442 of consensus"
7863..8160
/note="AlusC repeat: matches 1..295 of consensus"
8868..9402
/note="match: GSS: Em:AQ588534"
9425..9518
/note="MER3 repeat: matches 19..107 of consensus"
11033..11462
/note="TIGER2 repeat: matches 2267..2718 of consensus"
11450..11515
/note="MER8 repeat: matches 174..239 of consensus"
19576..19758
/note="LIM3A repeat: matches 5832..5999 of consensus"
19985..20012
/note="14 copies 2 mer ca 100 conserved"
20013..20062
/note="25 copies 2 mer ag 82 conserved"

```


REFERENCE AUTHORS TITLE JOURNAL REFERENCE AUTHORS

Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi;
Mammalia: Eutheria: Primates: Catarrhini: Homidae: Homo.
1 (bases 1 to 168957)
Birren, B., Linton, L., Nusbaum, C. and Lander, E.
Homo sapiens chromosome 12, clone RP11-206D3
Unpublished
2 (bases 1 to 168957)

Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F.,
Boguslavsky, L., Boukhgalter, B., Brown, A., Burkett, G.,
Boguslavsky, L., Castle, A., Choepel, Y., Colangelo, M., Collins, S.,
Collamore, A., Cooke, P., Deatellano, K., Dewar, K., Diaz, J.S.,
Dodge, S., Domingo, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D.,
Galagan, J., Gardina, S., Ginde, S., Goyette, M., Graham, L.,
Grand-pierre, N., Grant, G., Hagos, B., Heatford, A., Horton, L.,
Howard, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
Klein, J., Lacroque, K., Lamazares, R., Landers, T., Lebecky, J.,
Levine, R., Liu, C., Liu, G., Locke, K., MacDonald, P., Marquis, N.,
McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheters, R.,
Meldrum, J., Menus, L., Mihova, T., Miranda, C., Mlenga, V., Morrow, J.,
Murphy, T., Naylor, T.M., Norman, C.H., O'Connor, T., O'Donnell, P.,
O'Neill, D., Oliver, T.M., Oliver, J., Peterson, K., Pierre, N.,
Pisanu, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
Roy, A., Santos, R., Schauer, S., Severly, P., Spencer, B.,
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talmas, J.,
Tessfaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J.,
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,
Young, G., Zainoun, J., Zimmer, A. and Zody, M.

TITLE JOURNAL COMMENT

Direct Submission
Submitted (04-MAR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02144, USA
On May 25, 2000 this sequence version replaced g1:7705174.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIR

Web site: http://www.seq.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu

Project Information

Center project name: 206.D.3

Center clone name: 206.D.3

Summary Statistics

Sequencing vector: M13; W7815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 154625 bases at least Q40

Consensus quality: 161657 bases at least Q30

Consensus quality: 164463 bases at least Q20

Insert size: 166000; agarose-1p

Insert size: 166457; sum-of-coverage

Quality coverage: 4.0 in Q20 bases; sum-of-coverage

Quality coverage: 4.0 in Q20 bases; sum-of-coverage

NOTE: This is a 'working draft' sequence. It currently
consists of 26 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.

1181: contig of 1181 bp in length
1182: gap of 100 bp
1282: contig of 639 bp in length
1921: gap of 100 bp
2021: contig of 2460 bp in length
4481: gap of 100 bp
4481: 4580: gap of 100 bp
4581: 6784: contig of 2204 bp in length
6785: gap of 100 bp
6885: 9111: contig of 2227 bp in length
9112: gap of 100 bp

FEATURES

Source

9212 11337: contig of 2126 bp in length
11338 11437: gap of 100 bp
11438 14686: contig of 3249 bp in length
14687 14786: gap of 100 bp
14787 17181: contig of 2395 bp in length
17182 17281: gap of 100 bp
17282 20045: contig of 2764 bp in length
20046 20145: gap of 100 bp
20146 24869: contig of 4724 bp in length
24870 24969: gap of 100 bp
24970 30566: contig of 5597 bp in length
30567 30666: gap of 100 bp
30667 35562: contig of 4896 bp in length
35563 35662: gap of 100 bp
35663 43677: contig of 8015 bp in length
43678 43777: gap of 100 bp
43778 48684: contig of 4907 bp in length
48685 48784: gap of 100 bp
48785 55249: contig of 6465 bp in length
55250 55349: gap of 100 bp
55350 62047: contig of 6698 bp in length
62048 62147: gap of 100 bp
62148 70929: contig of 8782 bp in length
70930 71029: gap of 100 bp
71030 77706: contig of 6677 bp in length
77707 77806: gap of 100 bp
77807 84571: contig of 6765 bp in length
84572 84671: gap of 100 bp
84672 93794: contig of 9123 bp in length
93795 93894: gap of 100 bp
93895 102806: contig of 8912 bp in length
102807 102906: gap of 100 bp
102907 111867: contig of 8961 bp in length
111868 111967: gap of 100 bp
111969 123713: contig of 11746 bp in length
123714 123813: gap of 100 bp
123814 134343: contig of 10530 bp in length
134344 134443: gap of 100 bp
134444 150183: contig of 15740 bp in length
150184 150283: gap of 100 bp
150284 168957: contig of 18674 bp in length.

Location/Qualifiers

1. 168957
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="12"
/map="12"
/clone="RP11-206D3"
/clone_1b="RP11 Human Male BAC"
1. 1181
/note="assembly_fragment"
1282. 1920
/note="assembly_fragment"
clone_end:17
vector_side:right
2021. 4480
/note="assembly_fragment"
4581. 6784
/note="assembly_fragment"
6885. 9111
/note="assembly_fragment"
9212. 11337
/note="assembly_fragment"
11438. 14686
/note="assembly_fragment"
14787. 17181
/note="assembly_fragment"
17282. 20045
/note="assembly_fragment"
20146. 24869
/note="assembly_fragment"
24970. 30566
/note="assembly_fragment"


```

repeat_region /note="AluYb repeat: matches 199. .290 of consensus"
6913. .7099 /note="MLTIF repeat: matches 303. .477 of consensus"
repeat_region /note="AluY repeat: matches 1. .294 of consensus"
7100. .7381 /note="AluY repeat: matches 1. .294 of consensus"
repeat_region /note="MLTIF repeat: matches 477. .536 of consensus"
7382. .7439 /note="MLTIF repeat: matches 477. .536 of consensus"
misc_feature /note="match: GSS: Em:AQ21793"
7451. .7685 /note="match: GSS: Em:AQ21793"
repeat_region /note="L1R8 repeat: matches 446. .691 of consensus"
7493. .7693 /note="match: GSS: Em:AQ309297"
misc_feature /note="match: GSS: Em:AQ39210"
7498. .7688 /note="match: GSS: Em:AQ39210"
misc_feature /note="match: GSS: Em:AQ39210"
7509. .7678 /note="match: GSS: Em:AQ39210"
misc_feature /note="match: GSS: Em:AQ118984 Em:AQ118987 Em:AQ704859"
7625. .7810 /note="match: GSS: Em:AQ118984 Em:AQ118987 Em:AQ704859"
misc_feature /note="match: GSS: Em:AQ35679"
7686. .7837 /note="match: GSS: Em:AQ35679"
misc_feature /note="match: GSS: Em:AQ665254"
7686. .7837 /note="match: GSS: Em:AQ665254"
misc_feature /note="match: GSS: Em:AQ205158"
7687. .7859 /note="match: GSS: Em:AQ205158"
misc_feature /note="match: GSS: Em:AQ582263"
7687. .7859 /note="match: GSS: Em:AQ582263"
misc_feature /note="match: GSS: Em:AQ352039"
7687. .7859 /note="match: GSS: Em:AQ352039"
misc_feature /note="match: GSS: Em:AQ836127"
7687. .7859 /note="match: GSS: Em:AQ836127"
misc_feature /note="match: GSS: Em:AQ466383"
7687. .7859 /note="match: GSS: Em:AQ466383"
repeat_region /note="L1R8 repeat: matches 1. .194 of consensus"
8179. .8481 /note="L1R8 repeat: matches 1. .194 of consensus"
repeat_region /note="AluYb repeat: matches 1. .307 of consensus"
8487. .10618 /note="AluYb repeat: matches 1. .307 of consensus"
repeat_region /note="L1P repeat: matches 13. .2143 of consensus"
10619. .10909 /note="L1P repeat: matches 13. .2143 of consensus"
repeat_region /note="AluY repeat: matches 1. .291 of consensus"
10910. .13979 /note="AluY repeat: matches 1. .291 of consensus"
repeat_region /note="L1P repeat: matches 2143. .5220 of consensus"
13980. .14214 /note="L1P repeat: matches 2143. .5220 of consensus"
repeat_region /note="AluX repeat: matches 3. .237 of consensus"
14320. .14354 /note="AluX repeat: matches 3. .237 of consensus"
repeat_region /note="L1P repeat: matches 4963. .4997 of consensus"
14467. .14670 /note="L1P repeat: matches 4963. .4997 of consensus"
repeat_region /note="MER30 repeat: matches 1. .230 of consensus"
14845. .15243 /note="MER30 repeat: matches 1. .230 of consensus"
repeat_region /note="MLTIF repeat: matches 1. .493 of consensus"
14845. .15243 /note="MLTIF repeat: matches 1. .493 of consensus"
misc_feature /note="match: GSS: Em:AQ746910"
16403. .16476 /note="match: GSS: Em:AQ746910"
repeat_region /note="MIR repeat: matches 103. .175 of consensus"
16550. .16695 /note="MIR repeat: matches 103. .175 of consensus"
repeat_region /note="MIR repeat: matches 195. .240 of consensus"
16696. .16994 /note="MIR repeat: matches 195. .240 of consensus"
repeat_region /note="AluX repeat: matches 1. .305 of consensus"
16995. .17142 /note="AluX repeat: matches 1. .305 of consensus"
repeat_region /note="MIR repeat: matches 47. .195 of consensus"
17619. .17658 /note="MIR repeat: matches 47. .195 of consensus"
misc_feature /note="match: GSS: Em:AQ813272"
17619. .17658 /note="match: GSS: Em:AQ813272"
repeat_region /note="20 copies 2 mer ca 85% conserved"
17779. .18601 /note="20 copies 2 mer ca 85% conserved"
repeat_region /note="L2 repeat: matches 1720. .2710 of consensus"
18676. .18832 /note="L2 repeat: matches 1720. .2710 of consensus"
repeat_region /note="MIR repeat: matches 105. .262 of consensus"
18818. .18865 /note="MIR repeat: matches 105. .262 of consensus"
repeat_region /note="MIR repeat: matches 160. .207 of consensus"
18851. .18983 /note="MIR repeat: matches 160. .207 of consensus"
repeat_region /note="MIR repeat: matches 2. .139 of consensus"
19064. .19190 /note="MIR repeat: matches 2. .139 of consensus"
repeat_region /note="AluX repeat: matches 2. .128 of consensus"
19064. .19190 /note="AluX repeat: matches 2. .128 of consensus"

repeat_region 19191. .19436 /note="123 copies 2 mer ta 76% conserved"
repeat_region 19439. .19609 /note="AluX repeat: matches 135. .305 of consensus"
repeat_region 19921. .19956 /note="L2 repeat: matches 2587. .2620 of consensus"
repeat_region 19957. .20643 /note="MER67C repeat: matches 1. .710 of consensus"
repeat_region 20644. .20762 /note="L2 repeat: matches 2620. .2750 of consensus"
repeat_region 21182. .21263 /note="L2 repeat: matches 2620. .2750 of consensus"
repeat_region 21264. .21601 /note="41 copies 2 mer tt 78% conserved"
repeat_region 21794. .22002 /note="AluX repeat: matches 3. .312 of consensus"
repeat_region 22104. .22335 /note="MIR repeat: matches 16. .226 of consensus"
repeat_region 22605. .22646 /note="MIR repeat: matches 14. .251 of consensus"
repeat_region 22903. .23100 /note="L2 repeat: matches 2701. .2742 of consensus"
repeat_region 23147. .23188 /note="MIR repeat: matches 14. .219 of consensus"
repeat_region 23480. .23761 /note="L2 repeat: matches 2709. .2750 of consensus"
repeat_region 23771. .24252 /note="AluX repeat: matches 1. .280 of consensus"
repeat_region 25762. .25829 /note="MLTIF repeat: matches 3. .502 of consensus"
repeat_region 25830. .26119 /note="L1M4 repeat: matches 5161. .5222 of consensus"
repeat_region /note="AluX repeat: matches 1. .290 of consensus"

Query Match 15.4% Score 61.8; DB 9; Length 135693;
Best Local Similarity 85.2% Pred. No. 1e-07; 12; Indels 0; Gaps 0;
Matches 69; Conservative 0; Mismatches 12;

QY 322 agaatgggctctgtatgtcccaagaatggaagttagtggctatcataggaatgat 381
DB 99178 AGAGACAGGGCTCTGCTATGTTCGCCGCTGGATATGCACTGCTATTCACAGCATGAT 99119
DB 99118 CATGTGCACTACAGCCTTAA 99098

Search completed: January 31, 2002, 19:48:25
Job time: 1923 sec

```

us-09-641-831-5.rge

Fri Feb 1 08:56:33 2002

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 31, 2002, 19:50:07 ; Search time 67.87 Seconds
(without alignments)
1341.449 Million cell updates/sec

Title: US-09-641-831-5

Perfect score: 402

Sequence: 1 atgatgagagaccactgaaga.....atcatgactgcagccttga 402

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 351203 seqs, 11323899 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_NA:*
1: /cgn2_6/prodata/2/lna/5A.COMB.seq:*
2: /cgn2_6/prodata/2/lna/5B.COMB.seq:*
3: /cgn2_6/prodata/2/lna/5A.COMB.seq:*
4: /cgn2_6/prodata/2/lna/5B.COMB.seq:*
5: /cgn2_6/prodata/2/lna/PCTUS.COMB.seq:*
6: /cgn2_6/prodata/2/lna/Backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	48.4	12.0	87350	3	US-08-781-891-79
2	47.7	11.7	8353	3	US-08-611-587-1
3	46.6	11.6	2167	3	US-08-884-324-7
4	46.6	11.6	28994	3	US-08-884-324-14
5	44.4	11.0	3952	2	US-08-381-691-16
6	41	10.2	35060	3	US-08-814-095-7
7	39.2	9.8	7218	1	US-08-232-463-14
8	36.4	9.1	1334	3	US-08-884-324-8
9	36.4	9.1	11464	3	US-08-884-324-13
10	36.2	9.0	1696	4	US-09-124-238A-2
11	35.6	8.9	356	4	US-09-328-111-339
12	34.6	8.6	717	4	US-09-124-238A-7
13	34.6	8.6	1620	4	US-09-124-238A-32
14	34.6	8.6	1644	4	US-09-124-238A-9
15	34.6	8.6	1665	4	US-09-124-238A-33
16	34.6	8.6	1689	4	US-09-124-238A-22
17	34.4	8.6	5095	1	US-08-092-817-3
18	33.4	8.3	1597	1	US-08-166-316-1
19	33.2	8.3	1336	2	US-08-739-485-6
20	33	8.2	765	4	US-09-124-238A-21
21	32.4	8.1	1960	2	US-08-533-306A-1
22	32.4	8.1	1960	2	US-08-742-923A-1
23	32.4	8.1	2680	2	US-08-533-306A-5
24	32.4	8.1	2680	2	US-08-742-923A-5
25	32.4	8.1	2887	2	US-08-533-306A-3
26	32.4	8.1	2887	2	US-08-742-923A-3
27	32.2	8.0	4403765	4	US-09-103-840A-2

c	28	32.2	8.0	4411529	4	US-09-103-840A-1	Sequence 1, Appl
	29	32	8.0	2162	3	US-08-948-705-5	Sequence 5, Appl
	30	32	8.0	2236	3	US-08-948-705-4	Sequence 4, Appl
	31	32	8.0	2289	3	US-08-948-705-8	Sequence 8, Appl
	32	31.8	7.9	1782	1	US-08-374-155A-13	Sequence 13, Appl
	33	31.8	7.9	1782	1	US-08-785-396-13	Sequence 13, Appl
	34	31.4	7.8	22481	5	PCT-US95-07201-43	Sequence 43, Appl
	35	31.2	7.8	646	1	US-08-370-648-1	Sequence 1, Appl
	36	31.2	7.8	646	2	US-08-531-662B-1	Sequence 1, Appl
	37	31.2	7.8	646	3	US-08-669-161A-1	Sequence 1, Appl
	38	31.2	7.8	646	3	US-08-602-039-1	Sequence 1, Appl
	39	31.2	7.8	648	1	US-08-250-162A-1	Sequence 1, Appl
	40	31.2	7.8	20303	1	US-08-370-975B-6	Sequence 6, Appl
	41	31.2	7.8	26764	1	US-08-370-975B-1	Sequence 1, Appl
	42	31	7.7	624	4	US-09-385-982-414	Sequence 414, App
	43	31	7.7	2741	1	US-08-832-877-59	Sequence 59, Appl
	44	31	7.7	2741	2	US-08-832-877-59	Sequence 59, Appl
	45	30.8	7.7	1796	1	US-07-816-283-11	Sequence 11, Appl

ALIGNMENTS

RESULT 1
US-08-781-891-79
Sequence 79, Application US/08781891
Patent No. 6090620
GENERAL INFORMATION:
APPLICANT: Fu, Ying-Hui
APPLICANT: Yu, Chang-En
APPLICANT: Oshima, Junko
APPLICANT: Mulligan, John T.
APPLICANT: Schellenberg, Gerald D.
TITLE OF INVENTION: GENE AND GENE PRODUCTS RELATED TO
NUMBER OF SEQUENCES: 209
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/781,891
FILING DATE: 27-DEC-1996
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: No. 6090620tendburg Ph.D., Carol
REGISTRATION NUMBER: 39,317
REFERENCE/DOCKET NUMBER: 240052.419
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 79:
SEQUENCE CHARACTERISTICS:
LENGTH: 87350 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: Linear
US-08-781-891-79

Query Match 12.0%, Score 48.4, DB 3, Length 87350;
Best Local Similarity 78.4%, Pred. No. 1.8e-05;
Matches 58, Conservative 0, Mismatches 16, Indels 0, Gaps 0;
Oy 339 gggctgtgtatgttgcacgaatggaagtgatgtgcattcatcatgagatgatacatc 388

Comp Genet

||||| 56638 GGGTCTGCTGTGTGCCAGGTGGAGTGCAGTGCATTACAGACATGATCATACCA 56697
||||| 389 cactgagccttga 402
||||| 56698 TGCATAGCCTTGA 56711

RESULT 2

US-08-611-587-1
; Sequence 1, Application US/08611587

; Patent No. 6150091

; GENERAL INFORMATION:

; APPLICANT: PANDOLFO, MASSIMO

; APPLICANT: MONTERMINI, LAURA

; APPLICANT: MOLTO, MARIA D.

; APPLICANT: Koenig, Michael

; APPLICANT: Campuzano, Victoria

; APPLICANT: Cossee, Mireille

; TITLE OF INVENTION: Direct Diagnosis of Friedreich Ataxia

; NUMBER OF SEQUENCES: 33

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Fulbright & Jaworski L.L.P. Patent Dept.

; STREET: 1301 McKinney, Suite 5100

; CITY: Houston

; STATE: Texas

; COUNTRY: U.S.

; ZIP: 77010

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/611,587

; FILING DATE: 03-MAR-1996

; CLASSIFICATION: 436

; ATTORNEY/AGENT INFORMATION:

; NAME: Brashears-Macatee, Sarah J.

; REGISTRATION NUMBER: 38,087

; REFERENCE/DOCKET NUMBER: D-5901

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 713-651-5620

; TELEFAX: 713-651-5246

; TELEX: 76-2829

; INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 8353 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

; HYPOTHETICAL: NO

; ANTI-SENSE: NO

; ORIGINAL SOURCE:

; ORGANISM: Homo sapien

; POSITION IN GENOME:

; CHROMOSOME/SEGMENT: 9q13

; UNITS: bp

US-08-611-587-1

Query Match

Best Local Similarity 11.7%; Score 47; DB 3; Length 8353;

Matches 56; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Db 332 tcttgctatttccaggaatgaaggtagtgcttattcattagggatgatcatcagcac 391

||||| 1543 TCTTGTATGATGGCGGGCTGGAAATGCTGTGCTATTACAGGCACATCATAGCTCAC 1602

||||| 392 tgcagccttga 402

||||| 1603 TGCAGCCTTGA 1613

RESULT 3

US-08-884-324-7/c

; Sequence 7, Application US/08884324

; Patent No. 6060283

; GENERAL INFORMATION:

; APPLICANT: Takanori OKURA

; APPLICANT: Kakuji TORIGOE

; APPLICANT: Masahi KURIMOTO

; TITLE OF INVENTION: GENOMIC DNA ENCODING A POLYPEPTIDE CAPABLE

; OF INDUCING THE PRODUCTION OF INTERFERON-

; NUMBER OF SEQUENCES: 35

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: BROWDY AND NEIMARK

; STREET: 419 Seventh Street, N.W., Suite 300

; CITY: Washington

; STATE: D.C.

; COUNTRY: USA

; ZIP: 20004

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/884,324

; FILING DATE:

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: JP 185,305/96

; FILING DATE: 27-JUN-1996

; ATTORNEY/AGENT INFORMATION:

; NAME: BROWDY, Roger L.

; REGISTRATION NUMBER: 25,618

; REFERENCE/DOCKET NUMBER: OKURA-1

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 202-628-5197

; TELEFAX: 202-737-3528

; INFORMATION FOR SEQ ID NO: 7:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 2167 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; MOLECULE TYPE: Genomic DNA

; ORIGINAL SOURCE:

; ORGANISM: human

; TISSUE TYPE: placenta

; FEATURE:

; NAME/KEY: exon + 3'UTR

; LOCATION: 1..2167

; IDENTIFICATION METHOD: E

US-08-884-324-7

Query Match

Best Local Similarity 11.6%; Score 46.6; DB 3; Length 2167;

Matches 58; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

QY 323 gaaatggggtcttgcctatgttgcacgaatggaaggtagtgcttattcattagggatgac 382

||||| 1717 GAGACGGGGTCTCACTTTGTTGGCTAGGCTCGAGTGCATTACAGGCAAGATC 1658

||||| 383 atcatgcactgcagcct 399

||||| 1657 ACATTGCACATATAGCCT 1641

RESULT 4

US-08-884-324-14/c

; Sequence 14, Application US/08884324

; Patent No. 6060283

```

? IDENTIFICATION METHOD: E
? NAME/KEY: mat peptide
? LOCATION: 26828..27046
? IDENTIFICATION METHOD: S
? NAME/KEY: 3 UTR
? LOCATION: 27047..28994
? IDENTIFICATION METHOD: E
?
US-08-884-324-14

Query Match          11.6%  Score 46.6;  DB 3;  Length 28994;
Best Local Similarity 75.3%  Pred. No. 4.2e-05;
Matches 58; Conservative 0; Mismatches 19; Indels 0; Gaps
?
OY 323 gaatgggggtctgtgctatgttgcgccagaatggaagtgatgtgctatccatgaagcatgac 382
    ||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 28544 GAGACGGGGGCTCTACTCTTTGTTGCTCTAGGCTCGAGTGCAGTGTCTATTTCACAGCCAGATC 28485
    ||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

OY 383 atcatgcagtcagcact 399
    | ||||| |||||
Db 28484 ACATTGCACCTATAGCCT 28468
    ||||| |||||

```

```

NAME/KEY: mat peptide
LOCATION: 26828..27046
IDENTIFICATION METHOD: S
NAME/KEY: 3' UTR
LOCATION: 27047..28994
IDENTIFICATION METHOD: E
US-08-884-324-14

Query Match 11.6%; Score 46.6; DB 3; Length 28994;
Best Local Similarity 75.3%; Pred. No. 4.2e-05;
Matches 55; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

OY 323 gaattgggtcttgcctatgttgcacgaatggaaggtagtcgttcattcataggaatc 392
||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 28544 GAGACGGGGTCTCATTGTGTGCTAGCCTCGAGTGCAGTGGCTATTACAGGCAAGATC 28485

OY 383 atcatgcactgcagcct 399
| ||||| |||||
Db 28484 ACATTGCACCTATAGCCT 28468

RESULT 5
US-08-381-691-16
; Sequence 16, Application US/08381691
; Patent No. 5852224
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Alpha-lac Albumin Gene Constructs
; NUMBER OF SEQUENCES: 17
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/381.691
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3952 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHEetical: NO
; ANTI-SENSE: NO
US-08-381-691-16

Query Match 11.0%; Score 44.4; DB 2; Length 3952;
Best Local Similarity 73.1%; Pred. No. 8.5e-05;
Matches 57; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

OY 322 agaaatgggctctgtcatgttgcacgaatggaaggtagtcattcataggaatc 381
||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3606 AGAATGGGGTCTCACATGTGTGCCAGTGTACAGGACAGAGGCTATTACAGGTGTGAG 3665

OY 382 catcatgcactgcagcct 399
|| ||||| |||||
Db 3666 CACAGCACACTACAGCCT 3683

RESULT 6
US-08-814-095-7/c
; Sequence 7, Application US/08814095
; Patent No. 6025183
; GENERAL INFORMATION:
; APPLICANT: Soreq, Hermona
; APPLICANT: Zakut, Haim
; APPLICANT: Shani, Moshe
; TITLE OF INVENTION: TRANSGENIC ANIMAL ASSAY SYSTEM FOR
; TITLE OF INVENTION: ANTI-CHOLINESTERASE SUBSTANCES

```

us-09-641-831-6.rai

Fri Feb -1 08:56:35 2002

NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: KOHN & ASSOCIATES
STREET: 30500 NO. 6025183thwestern Highway, Suite 410
CITY: Farmington Hills
STATE: Michigan
COUNTRY: U.S.
ZIP: 48334
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA: US/08/814,095
FILING DATE:
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Montgomery, Ilene N.
REGISTRATION NUMBER: 38,972
REFERENCE/DOCKET NUMBER: 2391.00066
TELECOMMUNICATION INFORMATION:
TELEPHONE: (248) 539-5050
TELEFAX: (248) 539-5055
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 35060 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "Cosmid including ACHE
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
POSITION IN GENOME:
CHROMOSOME/SEGMENT: 7q22
FEATURE:
NAME/KEY: promoter
LOCATION: 4089..22464
OTHER INFORMATION: /function= "ACHE Promotor"
FEATURE:
NAME/KEY: exon
LOCATION: 22465..22537
OTHER INFORMATION: /function= "non-translated"
OTHER INFORMATION: /gene= "ACHE"
FEATURE:
NAME/KEY: exon
LOCATION: 24090..25177
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /function= "(translation start:
OTHER INFORMATION: 24110)"
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /gene= "ACHE"
OTHER INFORMATION: /number= 2
FEATURE:
NAME/KEY: exon
LOCATION: 25524..26009
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /gene= "ACHE"
OTHER INFORMATION: /number= 3
FEATURE:
NAME/KEY: exon
LOCATION: 27005..27274
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /gene= "ACHE"
OTHER INFORMATION: /number= 4

FEATURE:
NAME/KEY: exon
LOCATION: 27255..28007
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /gene= "ACHE"
OTHER INFORMATION: /number= 5
FEATURE:
NAME/KEY: terminator
LOCATION: 27385..27387
FEATURE:
NAME/KEY: exon
LOCATION: 28008..28129
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /gene= "ACHE"
OTHER INFORMATION: /number= 6
FEATURE:
NAME/KEY: terminator
LOCATION: 28129..28131
FEATURE:
NAME/KEY: exon
LOCATION: complement (34528..34895)
OTHER INFORMATION: /function= "arsenite resistance
OTHER INFORMATION: /gene=
OTHER INFORMATION: /gene= "AR"
OTHER INFORMATION: /number= 1
FEATURE:
NAME/KEY: exon
LOCATION: complement (34092..34358)
OTHER INFORMATION: /gene= "AR"
OTHER INFORMATION: /number= 2
FEATURE:
NAME/KEY: exon
LOCATION: complement (33779..33963)
OTHER INFORMATION: /gene= "AR"
OTHER INFORMATION: /number= 3
FEATURE:
NAME/KEY: exon
LOCATION: complement (33493..33591)
OTHER INFORMATION: /gene= "AR"
OTHER INFORMATION: /number= 4
FEATURE:
NAME/KEY: exon
LOCATION: complement (33297..33408)
OTHER INFORMATION: /gene= "AR"
OTHER INFORMATION: /number= 5
FEATURE:
NAME/KEY: exon
LOCATION: complement (32959..33094)
OTHER INFORMATION: /gene= "AR"
OTHER INFORMATION: /number= 6
FEATURE:
NAME/KEY: exon
LOCATION: complement (32569..32628)
OTHER INFORMATION: /gene= "AR"
OTHER INFORMATION: /number= 7
FEATURE:
NAME/KEY: exon
LOCATION: complement (32386..32468)
OTHER INFORMATION: /gene= "AR"
OTHER INFORMATION: /number= 8
FEATURE:
NAME/KEY: exon
LOCATION: complement (31894..32080)
OTHER INFORMATION: /gene= "AR"
OTHER INFORMATION: /number= 9
FEATURE:
NAME/KEY: exon
LOCATION: complement (31363..31534)
OTHER INFORMATION: /gene= "AR"
OTHER INFORMATION: /number= 10
FEATURE:

```
NAME/KEY: exon
LOCATION: complement (31131..31284)
OTHER INFORMATION: /gene="AR"
OTHER INFORMATION: /number=11
FEATURE:
NAME/KEY: exon
LOCATION: complement (30816..31011)
OTHER INFORMATION: /gene="AR"
OTHER INFORMATION: /number=12
FEATURE:
NAME/KEY: exon
LOCATION: complement (30470..30626)
OTHER INFORMATION: /gene="AR"
OTHER INFORMATION: /number=13
FEATURE:
NAME/KEY: exon
LOCATION: complement (30187..30274)
OTHER INFORMATION: /gene="AR"
OTHER INFORMATION: /number=14
FEATURE:
NAME/KEY: exon
LOCATION: complement (29945..30073)
OTHER INFORMATION: /gene="AR"
OTHER INFORMATION: /number=15
FEATURE:
NAME/KEY: exon
LOCATION: complement (29664..29856)
OTHER INFORMATION: /gene="ARS"
OTHER INFORMATION: /number=16
US-08-814-095-7

Query Match
Best Local Similarity 10.2%; Score 41; DB 3; Length 35060;
Matches 56; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

QY 322 agaatgggtctgtctatgttcccaagaatggaagtagtgcattcatagcattgat 381
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 12050 AGAGAGGCGCTGTGCTATGTTGCCAGACTGAGTTCGCTATTTACAGGTTCAT 11991
QY 382 catcatgacatgcagccttga 402
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 11990 CCTACTACGTGATCTGCAGGGA 11970

RESULT 7
US-08-232-463-14/C
Sequence 14, Application US/08232463
GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHEIFLINGER, F.
APPLICANT: FALKNER, F. G.
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0239
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
```

```
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: pTZgpt-F1s
US-08-232-463-14
```

```
Query Match
Best Local Similarity 9.8%; Score 39.2; DB 1; Length 7218;
Matches 17; Conservative 210; Mismatches 173; Indels 0; Gaps 0;

QY 3 gatggggccactgaaagacttcacaaagctagtccacatlaactcaacagccac 62
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
DB 1431 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1372
QY 63 caaggaaagtacattatctggaagcattccctggaaggaagctccctgggtttac 122
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
DB 1371 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1312
QY 123 tctaaaggtgctcgagcagcaggaaccattatcatctctaagtcgagaaggagg 182
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
DB 1311 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1252
QY 183 caaagcagaccctcgaagctccaactgcagctggaatgagttgtgcacatcatga 242
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
DB 1251 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1192
QY 243 gttgactctgagcagctccagaaggaagcagttccctcgtgtgaaggaatccctaagac 302
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
DB 1191 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1132
QY 303 cctcagctggtgtagtgcagaaatgggtctgtctatgtgccagaaatggaagtgt 362
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
DB 1131 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1072
QY 363 ggctattcatagcatgacatcagcagcagccttga 402
    ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1071 RRRRRATCCAGCTCCCTCGACCTGCACTGCAACCAAGCTCGGA 1032

RESULT 8
US-08-884-324-8/C
Sequence 8, Application US/08884324
GENERAL INFORMATION:
APPLICANT: Takano, I. OKURA
APPLICANT: Kakuji TORIGOE
APPLICANT: Masahiro KURIMOTO
TITLE OF INVENTION: GENOMIC DNA ENCODING A POLYPEPTIDE CAPABLE
NUMBER OF SEQUENCES: 35
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROWDY AND NEIMARK
STREET: 419 Seventh Street, N.W., Suite 300
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20004
```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model
Run on: January 31, 2002, 20:56:12 ; Search time 56.12 seconds
(without alignments)
175.548 Million cell updates/sec

Title: US-09-641-831-6
Perfect score: 695
Sequence: 1 MWRTEFHFKPSATLSNTA.....CCPEKVKVAIHRDHHALQP.133

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues
Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_1101.*
1: /SIDS2/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	695	100.0	133	22 AAU00033	Novel human protei
2	541	77.8	108	22 AAU00032	Novel human protei
3	541	77.8	190	22 AAU00031	Novel human protei
4	131.5	18.9	913	18 AAW12522	Protein tyrosine p
5	131.5	18.9	913	20 AAY25156	Human PTPH1 protei
6	128.5	18.5	150	21 AAB43364	Human ORFX ORF128
7	128	18.4	403	20 AAY07095	Colon cancer assoc
8	128	18.4	481	20 AAY07098	Colon cancer assoc
9	128	18.4	521	20 AAY07099	Colon cancer assoc
10	128	18.4	652	20 AAY07094	Colon cancer assoc
11	128	18.4	652	20 AAY07050	Renal cancer assoc

12	127.5	18.3	406	19 AAW70326	Secreted protein D
13	124.5	17.9	207	22 AAE03655	Human extracellular
14	122.5	17.6	86	22 AAB55840	PDZ encoded domain
15	122.5	17.6	86	22 AAB57831	Human LIM protein
16	122.5	17.6	86	22 AAB58041	Human LIM protein
17	121.5	17.5	127	20 AAY12510	Human 5' EST secre
18	121	17.4	181	22 AAG64378	Human PDZ protein
19	117.5	16.9	316	19 AAW58397	Homo sapiens HLIM-
20	117.5	16.9	329	19 AAW58396	Homo sapiens HLIM-
21	117.5	16.9	339	21 AAB56930	Human prostate can
22	117	16.8	37	22 AAB32879	Peptide #6916 enco
23	116.5	16.8	113	21 AAG02039	Human secreted pro
24	115.5	16.6	197	22 AAB93600	Human protein sequ
25	115.5	16.6	352	22 AAB95432	Human protein sequ
26	112.5	16.2	233	20 AAY29978	Human cell junctio
27	112.5	16.2	233	22 AAU07127	Human cell junctio
28	108.5	15.6	1112	21 AAB03833	Activin receptor b
29	108.5	15.6	1161	21 AAB03832	Rat synaptic scaff
30	108.5	15.6	1277	21 AAB03149	PDZ encoded domain
31	108	15.5	97	22 AAB55836	Tax interaction pr
32	108	15.5	97	22 AAB57627	Tax interaction pr
33	108	15.5	95	21 AAG02034	Human secreted pro
34	107.5	15.5	153	21 AAB36475	Human LMP-3 (HIMP-
35	107.5	15.5	223	20 AAB7845	Human truncated bo
36	107.5	15.5	223	21 AAB36473	Human LMP-1 amino
37	107.5	15.5	423	21 AAB36474	Human LMP-2 (HIMP-
38	107.5	15.5	423	21 AAB36474	Rat bone mineralis
39	107.5	15.5	457	20 AAW97843	Human bone mineral
40	107.5	15.5	457	20 AAW97844	Rat LIM mineralisa
41	107.5	15.5	457	21 AAB36471	Human LIM minerali
42	107.5	15.5	457	21 AAB36472	Protein containing
43	104.5	15.0	80	21 AAB53231	PDZ encoded domain
44	104.5	15.0	344	22 AAB55834	Human post-synapti
45	104.5	15.0	344	22 AAB57623	

ALIGNMENTS

RESULT	1				
AAU00033					
ID	AAU00033	standard; Protein; 133 AA.			
XX	AAU00033;				
AC	AAU00033;				
XX	11-MAY-2001 (first entry)				
DT	11-MAY-2001 (first entry)				
XX	Novel human protein, NHP#3.				
DE	Novel human protein, NHP#3.				
XX	Human; novel human protein; NHP#3; gene therapy; drug screening;				
KW	obesity; high blood pressure.				
KW	obesity; high blood pressure.				
XX	Homo sapiens.				
OS	Homo sapiens.				
XX	Key	Location/Qualifiers			
PH	Misc_feature	1			
FT		/note= "Absent if the true start codon is			
FT		nucleotides 4-6 of sequence in AAS00041, the authors			
FT		are uncertain which is the true start codon"			
XX	WO200114422-A1.				
PN	01-MAR-2001.				
XX	18-AUG-2000; 2000WO-US22815.				
XX	24-AUG-1999; 99US-0150511.				
PR	(LEXI-) LEXICON GENETICS INC.				
XX	Turner CA, Zambrowicz B, Friedrich G, Nehls M, Sands AT;				
XX	WPI; 2001-218430/22.				
DR					

DR N-PSDB; AAS00031.
XX Novel human polynucleotides isolated from human mammary gland cDNA
PT library, encodes novel human proteins which are useful in diagnosis,
PT drug screening, clinical trial monitoring or treating behavioural
PT disorders
XX
PS Claim 2; Page 24-25; 29pp; English.
XX
CC The sequence is a novel human protein, NHP#1, which shares
CC structural motifs with human APXL protein. Nucleotide constructs
CC encoding functional NHPs are used in gene therapy approaches for the
CC modulation of NHP expression. NHP oligonucleotides can be used as
CC hybridisation probes for screening libraries and assessing NHP gene
CC expression patterns. Also, labeled NHP nucleotide probes can be used to
CC screen a human genomic library. The NHP nucleotide sequences are also
CC useful in drug screening techniques for treating symptomatic or
CC phenotypic manifestations of perturbing the normal function of NHP in the
CC body. Examples of such manifestations may include obesity and high blood
CC pressure.
XX
SQ Sequence 190 AA;

Query Match 77.8%; Score 541; DB 22; Length 190;
Best Local Similarity 100.0%; Pred. No. 1.5e-57;
Matches 107; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MMRTEDFHPSATLNSNTATKGRYIYLEAFLEGGAPWGFTLKGLEHGEPLIISKVEEG 60
DB 1 mrrttedfhpsatlnsntatkgryiyleaflegapwgftlkgglehgeplliiskveeg 60
QY 61 GKADTLSSKLQAGDEVHINEVTLSSRKEAVSLVKGSKYKTLRLVVR 107
DB 61 gkadtlssklqagdevhinevtlssrkeavslvkgskytllrlvvr 107

RESULT 4
AAW12522
ID AAW12522 standard; Protein; 913 AA.
XX
AC AAW12522;
XX
XX 22-MAY-1997 (first entry)
DE Protein tyrosine phosphatase that localises to focal adhesion.
XX
XX Protein tyrosine phosphatase; PTPH1; focal adhesion;
KW protein tyrosine kinase; malignancy; cancer; gene therapy;
KW retrovirus; vector.
XX
OS Homo sapiens.
XX
XX Key Location/Qualifiers
FH 30..357
FT Region /note= "N-terminal region shows homology with the
FT /note= "N-terminal regions of the talin family"
FT Region 357..663
FT /note= "central region includes sequences with
FT features of sites of phosphorylation by
FT casein kinase II and p34cdc2"
FT Modified-site 372
FT /label= Phosphorylation
FT /note= "putative p34cdc2 phosphorylation site"
FT Modified-site 381
FT /label= Phosphorylation
FT /note= "putative p34cdc2 phosphorylation site"
FT Modified-site 424..428
FT /label= Phosphorylation
FT /note= "putative casein kinase II phosphorylation
FT site"
FT Modified-site 438..442
FT /label= Phosphorylation

FT /note= "putative casein kinase II phosphorylation
FT site"
FT Modified-site 489..492
FT /label= Phosphorylation
FT /note= "putative casein kinase II phosphorylation
FT site"
FT Modified-site 514..517
FT /label= Phosphorylation
FT /note= "putative casein kinase II phosphorylation
FT site"
FT Modified-site 543..547
FT /label= Phosphorylation
FT /note= "putative casein kinase II phosphorylation
FT site"
FT Modified-site 607..610
FT /label= Phosphorylation
FT /note= "putative casein kinase II phosphorylation
FT site"
FT Region 664..931
FT /note= "c-terminal region shows homology to "
XX
XX US5595911-A.
XX
XX 21-JAN-1997.
XX
XX 14-MAR-1990; 90US-0494036.
XX
XX 01-MAR-1991; 91US-0663579.
XX
XX 14-MAR-1990; 90US-0494036.
XX
XX 16-AUG-1993; 93US-0107420.
XX
XX (COLD-) COLD SPRING HARBOR LAB.
XX
XX Tonks NK;
XX
XX WPI; 1997-107583/10.
XX
XX N-PSDB; AAT58627.
XX
XX DNA encoding protein tyrosine phosphatase - for gene therapy of
XX cancer
XX
XX Disclosure; Fig 1A-B; 12pp; English.
XX
XX A protein tyrosine phosphatase (PTPH1) (AAW12522) catalyses the
XX dephosphorylation of proteins in which tyrosyl residues have been
XX phosphorylated through the action of a protein tyrosine kinase
XX (PTK). It localises to focal adhesions, a major site of action of
XX oncogenic PTKs. PTPH1 is the product of a cDNA clone (AAT58627)
XX obtd. from HeLa cells. The PTPH1 cDNA can be incorporated into a
XX vector (pref. retroviral) to allow expression of PTPH1 in mammalian
XX in sufficient quantities to overcome or counteract PTK activity.
XX Phosphorylation of tyrosine residues at abnormal levels is
XX prevented or reversed, resulting in the prevention or reversal of
XX malignancy of cells.
XX
SQ Sequence 913 AA;

Query Match 18.9%; Score 131.5; DB 18; Length 913;
Best Local Similarity 28.5%; Pred. No. 9e-07;
Matches 35; Conservative 23; Mismatches 46; Indels 19; Gaps 4;
QY 6 EDFHPSATLNSNTATKGRY-----IYLEAFLEGGAPWGFTLKGLEHGEPLI 53
DB 479 ddfrh-vtkggstedsaggyckndngdsylvliripdedgkfglnkgvdkmplv 536
QY 54 ISKVEEGKADTLSSKLQAGDEVHINEVTLSS-SRKEAVSLVKGSKYKT----LRLVVR 108
DB 537 vsrlnpespadtclpknegdqivlndgrdischdqvvmfikasreshrelalvrr 596
QY 109 NGV 111
DB 597 rav 599

RESULT 5
 ID AAY25156 standard; Protein: 913 AA.
 AC AAY25156.
 DT 07-SEP-1999 (first entry)
 DE Human PTPH1 protein.
 KW PTPH1; human; protein tyrosine phosphatase; focal adhesion; cancer;
 KW localisation; treatment; overexpression; oncogenic; cell transformation;
 KW prevention; phosphotyrosine; disease; malignant.
 OS Homo sapiens.
 FH Key
 FT Region
 FT /note= "region of homology to the N-terminal domain
 FT of band 4.1, ezrin and talin. This region is
 FT known to be important for localisation to focal
 FT adhesions"
 FT Modified-site 372
 FT /note= "potential p34cdc2 phosphorylation site"
 FT Modified-site 381
 FT /note= "potential p34cdc2 phosphorylation site"
 FT Modified-site 424.428
 FT /note= "potential casein kinase II phosphorylation site"
 FT Modified-site 434
 FT /note= "potential p34cdc2 phosphorylation site"
 FT Modified-site 438.442
 FT /note= "potential casein kinase II phosphorylation site"
 FT Modified-site 489.492
 FT /note= "potential casein kinase II phosphorylation site"
 FT Modified-site 514.518
 FT /note= "potential casein kinase II phosphorylation site"
 FT Modified-site 543.547
 FT /note= "potential casein kinase II phosphorylation site"
 FT Modified-site 607.610
 FT /note= "potential casein kinase II phosphorylation site"
 FT Modified-site
 FT US563781-A.
 PN 26-JAN-1999.
 PD 04-DEC-1996; 96US-0759536.
 XX 01-MAR-1991; 91US-0663579.
 PR 14-MAR-1990; 90US-0494036.
 PR 16-AUG-1993; 93US-0107420.
 PR 04-DEC-1996; 96US-0759536.
 PA (COLD-) COLD SPRING HARBOR LAB.
 XX
 PI Tonks NK;
 XX WPI; 1999-131308/11.
 DR N-PSDB; AAY78463.
 XX
 PT Protein tyrosine phosphatase PTPH1 - encoded by DNA of Hela cells
 PS Claim 1; Fig 1A-B; 12pp; English.
 CC This sequence represents a novel protein tyrosine phosphatase, PTPH1,
 CC isolated from Hela cells. The protein of the invention appears to
 CC localise to focal adhesions and is therefore potentially useful in the
 CC treatment of cancer. Overexpression of PTPH1 can be used to counter the
 CC effects of oncogenic protein tyrosine kinases such as those of
 CC transforming viruses and for interfering with or reversing cell
 CC transformation. This would provide a means of preventing or reversing
 CC abnormally high levels of phosphotyrosine associated with any disease or

CC condition such as preventing or reversing malignancy associated with the
 CC activity of a protein tyrosine kinase.
 XX
 SQ Sequence 913 AA;
 Query Match 18.9%; Score 131.5; DB 20; Length 913;
 Best Local Similarity 28.5%; Pred. No. 9e-07;
 Matches 35; Conservative 23; Mismatches 46; Indels 19; Gaps 4;
 QY 6 EDFHKEPATLNSMTATGR-----IYLEAFLEGAPWFTLGGLEHGPPLI 53
 DB 479 ddfhr--vfkgsdedsqycdkndngdsylvilitpedgkfgfnlkygdqkmpiv 536
 QY 54 ISKVEEGGRADTLSSKQADDEVVHNEVTLSS-SREAVSLVKGSKT-----LRLVRR 108
 DB 537 vsrlmpespadctqlprlnegqdlvllngldsenhthqvmfllkasreshrelalvrr 596
 QY 109 NGV 111
 DB 597 rav 599
 RESULT 6
 ID AAB43364 standard; Protein: 150 AA.
 AC AAB43364;
 DT 08-FEB-2001 (first entry)
 DE Human ORFX ORF3128 polypeptide sequence SEQ ID NO:6256.
 XX
 XX Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
 XX vulnery; antiporiatic; antiparkinsonian; neurotropic; neuroprotective;
 XX anticonvulsant; osteopathic; antirheumatic; immunosuppressant; cardiac;
 XX immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
 XX antitvital; antibacterial; antitumoral; antineoplastic; antihypertensive;
 XX antianemic; gene therapy; cancer; proliferative disorder; hypertension;
 XX neurodegenerative disorder; osteoarthritis; graft vs host disease;
 XX cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 XX cholesterol ester storage; systemic lupus erythematosus; infection;
 XX severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 XX allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 XX bone damage; cartilage damage; antiinflammatory disease; coagulation;
 XX thrombosis; contraceptive.
 OS Homo sapiens.
 XX
 XX WO200058473-A2.
 PN 05-OCT-2000.
 PD 31-MAR-2000; 2000WO-US08621.
 XX
 XX 31-MAR-1999; 99US-0127607.
 PR 02-APR-1999; 99US-0127636.
 PR 05-APR-1999; 99US-0127728.
 PR 30-MAR-2000; 2000US-0540763.
 PA (CURA-) CURAGEN CORP.
 XX
 PI Shinkets RA, Leach M;
 XX
 XX WPI; 2000-602362/57.
 DR N-PSDB; AAC77573.
 XX
 PT Novel nucleic acids and peptides derived from open reading frame X,
 PT useful for treating e.g. cancers, proliferative disorders,
 PT neurodegenerative disorders and cardiovascular disease -
 PS Claim 11; Page 5443; 5507pp; English.

XX	15-JUL-1998;	98WO-US14679.	
PF			
XX	22-JUN-1998;	98US-0102322.	
XX	17-JUL-1997;	97US-0896164.	
PR	10-OCT-1997;	97US-0061599.	
PR	10-OCT-1997;	97US-0061765.	
PR	10-OCT-1997;	97US-0948705.	
PR	11-OCT-1997;	97GB-0021697.	
XX			
PA	(LUDW-) LUDWIG INST CANCER RES.		
XX			
PI	Chen Y, Gout I, Gure A, O'Hare M, Obata Y, Old L;		
PI	pfreundschuh M, Sahin U, Scanlan MJ, Stockert E;		
PI	Tureci O;		

XX	15-JUL-1998;	98WO-US14679.	
XX	XX		
XX	23-JUN-1998;	98US-0102322.	
PR	17-JUL-1997;	97US-0896164.	
PR	10-OCT-1997;	97US-0061599.	
PR	10-OCT-1997;	97US-0061785.	
PR	10-OCT-1997;	97US-0948705.	
PR	11-OCT-1997;	97GB-0021697.	
XX	(LUDW-) LUDWIG INST CANCER RES.		
XX	XX		
PI	Chen Y, Gout I, Gure A, O'Hare M, Obata Y, Old LJ;		
PI	Pfreundschuh M, Sahin U, Scanlan MJ, Stockert E;		

RESULT 14	RESULT 15
AAB55840	AAB57631
ID AAB55840 standard; Peptide; 86 AA.	ID AAB57631 standard; Protein; 86 AA.
XX AC AAB55840;	XX AC AAB57631;
XX DT 07-MAR-2001 (first entry)	XX DT 12-MAR-2001 (first entry)
XX DE PDZ encoded domain #10.	XX DE Human LIM protein PDZ domain.
XX KW Hematopoietic cell; PDZ; PL; autoimmune disease; inflammation;	XX KW Endothelial cell; haematopoietic cell; PDZ domain protein;
XX KW allergy; asthma; multiple sclerosis; cancer; infection.	XX KW PL domain protein; leukocyte activation; synapse formation;
XX OS Synthetic.	XX KW transmembrane neurotransmitter receptor; autoimmune disease;
XX PN WO200069896-A2.	XX KW transplantation rejection; inflammation; allergy;
XX PD 23-NOV-2000.	XX KW inflammatory bowel disease; ulcerative colitis; ileitis; psoriasis;
XX PF 12-MAY-2000; 2000WO-US13161.	XX KW asthma; atopic dermatitis; atherosclerosis; cancer; infectious disease;
XX PR 14-MAY-1999; 99US-0134114.	XX OS Homo sapiens.
XX PR 14-MAY-1999; 99US-0134117.	XX PN WO200069897-A2.
XX PR 14-MAY-1999; 99US-0134118.	XX PD 23-NOV-2000.
XX PR 21-OCT-1999; 99US-0160860.	XX PF 12-MAY-2000; 2000WO-US13166.
XX PR 29-OCT-1999; 99US-0162498.	XX PR 14-MAY-1999; 99US-0134114.
XX PR 13-DEC-1999; 99US-0170453.	XX PR 14-MAY-1999; 99US-0134117.
XX PR 14-JAN-2000; 2000US-0176193.	XX PR 14-MAY-1999; 99US-0134118.
XX PR 14-FEB-2000; 2000US-0182296.	XX PR 21-OCT-1999; 99US-0160860.
XX PR 11-APR-2000; 2000US-0196460.	XX PR 29-OCT-1999; 99US-0162498.
XX PR 11-APR-2000; 2000US-0196527.	XX PR 13-DEC-1999; 99US-0170453.
XX FA (ARBO-) ARBOR VITA CORP.	XX PR 14-JAN-2000; 2000US-0176193.
XX LU PS;	XX PR 14-FEB-2000; 2000US-0182296.
XX WPI; 2001-080245/09.	XX PR 11-APR-2000; 2000US-0196460.
XX DR	XX PR 11-APR-2000; 2000US-0196527.
XX PT Modulating a biological function of an endothelial cell or	XX PA (ARBO-) ARBOR VITA CORP.
XX PT hematopoietic cell, useful for treating autoimmune diseases and	XX PI Lu PS;
XX PT infectious diseases, by administering an antagonist that inhibits	XX PI WPI; 2001-025003/03.
XX PT binding between a PDZ protein and a PL protein -	XX DR New inhibitors of binding of a PDZ protein and PL protein for
XX PS Disclosure; Page 28-43; 141pp; English.	XX PS treating diseases characterized by inflammatory and humoral immune
XX CC The present invention relates to a new method for modulating a	XX CC responses, e.g. inflammation, cancer -
XX CC biological function of an endothelial cell or hematopoietic cell. The	XX CC Disclosure; Page 32; 139pp; English.
XX CC method involves introducing into a cell, an antagonist that inhibits	XX CC The present invention relates to a method for modulating a biological
XX CC binding between a PDZ protein and a PL protein. The inhibitor is used	XX CC function of an endothelial cell or hematopoietic cell, comprises
XX CC to treat a disease mediated by hematopoietic cells, e.g. autoimmune	XX CC introducing into a cell an antagonist that inhibits binding between a
XX CC disease. It may also be used to prevent transplantation rejection of	XX CC PDZ domain protein and a PL domain protein to result in inhibition of
XX CC a solid organ transplant. The method may also be used in the treatment	XX CC leukocyte activation. The present sequence is a PDZ domain. PDZ domains
XX CC of inflammation, allergy, inflammatory bowel diseases, ulcerative	XX CC of proteins are named after three prototypical proteins: PSD95,
XX CC colitis, ileitis, psoriasis, asthma, atopic dermatitis, autoimmune	XX CC drosophila large disc protein and Zonula Occludin 1 protein. PDZ domain
XX CC diseases (e.g. rheumatoid arthritis, multiple sclerosis,	XX CC proteins are involved in synapse formation by organising transmembrane
XX CC graft rejection, transplantation rejection), atherosclerosis, cancers,	XX CC neurotransmitter receptors through intracellular interactions. The
XX CC infectious diseases, ischemia, vasculitis and Crohn's disease.	XX CC inhibitors identified by the present invention can be used to treat a
XX SQ Sequence 86 AA;	XX CC disease mediated by hematopoietic cells, e.g. autoimmune disease,
	XX CC inflammation, allergy (e.g. drug allergies), inflammatory bowel diseases,
	XX CC ulcerative colitis, ileitis, psoriasis, respiratory allergic diseases
	XX CC (e.g. asthma), atopic dermatitis, autoimmune diseases (e.g. rheumatoid
	XX CC arthritis, multiple sclerosis, insulin-dependent diabetes, Hashimoto
	XX CC thyroiditis, osteoarthritis), atherosclerosis, cancers, infectious
	XX CC diseases (e.g. viral infection), ischemia, vasculitis and Crohn's
	XX CC disease. The inhibitors can also be used to prevent transplantation
	XX CC rejection of a solid organ transplant.
	XX SQ Sequence 86 AA;

Fri Feb 1 08:56:35 2002

us-09-641-831-6.rag

Page 10

	Query Match	17.6%	Score 122.5;	DB 22;	Length 86;
	Best Local Similarity	35.9%;	Pred. No. 3.6e-07;		
	Matches 28; Conservative	18;	Mismatches 29;	Indels 3;	Gaps 2;
Oy	32 LEGAPMGFTLGGLEHGEPLIISKVEEGGRADTLSSKLQAGDEVYVHNENVTLS-SRKE 90	: :	: :	:	
Dd	8 lvpapwffirldggxdfmplltissikdgkx--aqanvrlgvdvlslidlnagmthle 65	: :	: :	:	
Oy	91 AVSLVGSKYTLLRLVRR 108	: :	: :	:	
Dd	66 agnkikgcctgsimntlqr 83	: :	: :	:	

Search completed: January 31, 2002, 20:56:12
Job time: 80 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 20:57:10 ; Search time 29.66 Seconds
(without alignments)
100.908 Million cell updates/sec

Title: US-09-641-831-6
Perfect score: 695
Sequence: 1 MWRTTEDFHRPSATLNSNTA.....CCPEKVKVAIHRDHDLQ 133

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/6C_COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/6D_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution..

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	118.5	17.1	327	2	US-08-739-485-10
2	117.5	16.9	316	2	US-08-739-485-5
3	117.5	16.9	329	2	US-08-739-485-3
4	117	16.8	75	3	US-09-100-804-21
5	116.5	16.8	330	2	US-08-739-485-11
6	112.5	16.2	233	2	US-09-151-611-1
7	112.5	16.2	233	4	US-09-370-102-1
8	109	15.7	297	2	US-09-151-611-3
9	109	15.7	297	4	US-09-370-102-3
10	107.5	15.5	223	4	US-09-124-238A-34
11	107.5	15.5	457	4	US-09-124-238A-1
12	107.5	15.5	457	4	US-09-124-238A-10
13	104.5	15.0	454	1	US-08-166-316-2
14	103.5	14.9	73	3	US-09-100-804-28
15	97	14.0	182	3	US-09-045-632-29
16	97	14.0	283	3	US-09-045-632-27
17	97	14.0	498	3	US-09-045-632-30
18	97	14.0	502	3	US-09-045-632-24
19	97	14.0	541	3	US-09-045-632-36
20	97	14.0	599	3	US-09-045-632-28
21	97	14.0	602	3	US-09-045-632-20
22	97	14.0	642	3	US-09-045-632-35
23	97	14.0	702	3	US-09-045-632-15
24	97	14.0	818	3	US-09-045-632-25
25	97	14.0	861	3	US-09-045-632-34
26	97	14.0	918	3	US-09-045-632-21
27	97	14.0	961	3	US-09-045-632-33

28 97 14.0 1018 3 US-09-045-632-16 Sequence 16, Appl
29 97 14.0 1061 3 US-09-045-632-32 Sequence 32, Appl
30 97 14.0 1112 3 US-09-045-632-2 Sequence 2, Appl
31 97 14.0 1112 3 US-09-045-632-3 Sequence 3, Appl
32 95 13.7 505 1 US-08-123-161A-14 Sequence 14, Appl
33 95 13.7 505 1 US-08-483-278-14 Sequence 14, Appl
34 93.5 13.5 80 3 US-08-545-860D-54 Sequence 54, Appl
35 93.5 13.5 80 5 PCT-US94-04496-54 Sequence 54, Appl
36 92.5 13.3 73 3 US-09-100-804-25 Sequence 25, Appl
37 92.5 13.3 74 3 US-09-100-804-34 Sequence 34, Appl
38 90 12.9 82 3 US-09-045-632-9 Sequence 9, Appl
39 90 12.9 398 3 US-09-045-632-31 Sequence 31, Appl
40 90 12.9 441 3 US-09-045-632-37 Sequence 37, Appl
41 88.5 12.7 86 3 US-08-545-860D-53 Sequence 53, Appl
42 88.5 12.7 86 5 PCT-US94-04496-53 Sequence 53, Appl
43 87.5 12.6 75 3 US-09-100-804-22 Sequence 22, Appl
44 87.5 12.6 79 3 US-09-100-804-27 Sequence 27, Appl
45 85 12.2 2037 4 US-09-306-998-3 Sequence 3, Appl

ALIGNMENTS

RESULT 1
US-08-739-485-10
; Sequence 10, Application US/08739485
; Patent No. 5863898
; GENERAL INFORMATION:
; APPLICANT: Goli, Surya K.
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Bandman, Olga
; TITLE OF INVENTION: NOVEL HUMAN LIM PROTEINS
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: US
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/739,485
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0142 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 327 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 1020151
US-08-739-485-10

Query Match 17.1% Score 118.5 DB 2: Length 327;

us-09-641-831-6.ra1

Fri Feb. 1 08:56:35 2002

```

; TITLE OF INVENTION: EXPRESSION OF NUCLEOTIDE SEQUENCES FOR NOVEL PROTEIN
; TITLE OF INVENTION: TYROSINE PHOSPHATASES
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.
; STREET: 600 ATLANTIC AVENUE
; CITY: BOSTON
; STATE: MASSACHUSETTS
; COUNTRY: USA
; ZIP: 02210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/100,804
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/596,291
; FILING DATE: 09-AUG-1996
; APPLICATION NUMBER: US 08/115,573
; FILING DATE: 01-SEP-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/09943
; FILING DATE: 01-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: GATES, EDWARD R.
; REGISTRATION NUMBER: 31,616
; REFERENCE/DOCKET NUMBER: LO461/7003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-720-3500
; TELEFAX: 617-720-2441
; TELEX:
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 75 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; US-09-100-804-21

Query Match 16.8%; Score 117; DB 3; Length 75;
Best Local Similarity 38.8%; Pred. No. 2.3e-07;
Matches 26; Conservative 12; Mismatches 27; Indels 2; Gaps 2;

QY 33 EGGAPWGFTLKGLEHGEPLIISKVEEGKADTLSSKQLQAGDEVVHINEVTLSS-SRKEA 91
Db 2 EDGKP-GFNLKGVDQKNPLVVRINSPADTCIPKLNEDQIVLNGRDISEHEDQV 60

QY 92 VSLVKS 98
Db 61 VMFIKAS 67

RESULT 5
US-08-739-485-11
; Sequence 11, Application US/08739485
; Patent No. 5863898
; GENERAL INFORMATION:
; APPLICANT: Goli, Surya K.
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Bandman, Olga
; TITLE OF INVENTION: NOVEL HUMAN LIM PROTEINS
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive

```

```

; CITY: Palo Alto
; STATE: CA
; COUNTRY: US
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/739,485
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0142 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 330 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; ORIGINAL SOURCE:
; HAPLOTYPE: GenBank
; IMMEDIATE SOURCE:
; LIBRARY: 887580
; US-08-739-485-11

Query Match 16.8%; Score 116.5; DB 2; Length 330;
Best Local Similarity 41.0%; Pred. No. 2.4e-06;
Matches 32; Conservative 9; Mismatches 34; Indels 3; Gaps 2;

QY 32 LEGGAPWGFTLKGLEHGEPLIISKVEEGKADTLSSKQLQAGDEVVHIN-EVTLSSSRKE 90
Db 7 LRGPSPWGFRLVGRDFSPAPLTISRVAHGSKA--ALAAALCPGDSIQAINGESTELMTHLE 64

QY 91 AVSLVRGSKYKTLRLVVR 108
Db 65 AQNRKRGCHDHLTLVSVR 82

RESULT 6
US-09-151-611-1
; Sequence 1, Application US/09151611
; Patent No. 5958731
; GENERAL INFORMATION:
; APPLICANT: Yue, Henry
; APPLICANT: Au-Young, Janice
; APPLICANT: Patterson, Chandra
; TITLE OF INVENTION: CELL JUNCTION PDZ PROTEIN
; FILE REFERENCE: PF-0599 US
; CURRENT APPLICATION NUMBER: US/09/151,611
; CURRENT FILING DATE: 1998-09-11
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PERL Program
; SEQ ID NO 1
; LENGTH: 233
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: 1974337
; US-09-151-611-1

```



```

; PRIOR APPLICATION
; PRIOR FILING DATE
; NUMBER OF SEQ ID
; SOFTWARE: MS Word

```


Patent No. 6066472
GENERAL INFORMATION:
APPLICANT: GONEZ, LEONEL JORGE
APPLICANT: SARAS, JAN
APPLICANT: CLAESON-WEISH, LENA
APPLICANT: HELDIN, CARL-HENRIK
TITLE OF INVENTION: PRIMARY STRUCTURE AND FUNCTIONAL
TITLE OF INVENTION: EXPRESSION OF NUCLEOTIDE SEQUENCES FOR NOVEL PROTEIN
TITLE OF INVENTION: TYROSINE PHOSPHATASES
NUMBER OF SEQUENCES: 34
CORRESPONDENCE ADDRESS:
ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.
STREET: 600 ATLANTIC AVENUE
CITY: BOSTON
STATE: MASSACHUSETTS
COUNTRY: USA
ZIP: 02210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/100,804
FILING DATE:
CLASSIFICATION:
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/596,291
FILING DATE: 09-AUG-1996
APPLICATION NUMBER: US 08/115,573
FILING DATE: 01-SEP-1993
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/09943
FILING DATE: 01-SEP-1994
ATTORNEY/AGENT INFORMATION:
NAME: GATES, EDWARD R.
REGISTRATION NUMBER: 31,616
REFERENCE/DOCKET NUMBER: 10461/7003
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-720-3500
TELEFAX: 617-720-2441
TELEX:
INFORMATION FOR SEQ ID NO: 28:
SEQUENCE CHARACTERISTICS:
LENGTH: 73 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-09-100-804-28

Query Match 14.9%; Score 103.5; DB 3; Length 73;
Best Local Similarity 35.6%; Pred. No. 1.1e-05;
Matches 26; Conservative 19; Mismatches 25; Indels 3; Gaps 3;

QY 34 GGPMPGFTLKGGHEPLISVEEGKADTSSKLGADDEVVHINEVTL-SSSRKAV 92
DB 2 GSTGLGFNTVGG-EDGEGLIFSLAGPAD-LSGLRKGDOLSVNGVDLRNASHQAA 59
QY 93 SLVKSYSKTLRLV 105
DB 60 IALKNAGQTVTTI 72

RESULT 15
US-09-045-632-29
Sequence 29, Application US/09045632
Patent No. 6001575
GENERAL INFORMATION:
APPLICANT: HUGANIR, Richard L.

APPLICANT: Dong, Hualing
TITLE OF INVENTION: THERAPEUTIC USES OF GRIP AND
TITLE OF INVENTION: GRIP-RELATED MOLECULES
NUMBER OF SEQUENCES: 105
CORRESPONDENCE ADDRESS:
ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
STREET: 130 Water Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/045,632
FILING DATE: 19-MAR-1998
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 60/041,016
FILING DATE: 19-MAR-1997
ATTORNEY/AGENT INFORMATION:
NAME: Corless, Peter F.
REGISTRATION NUMBER: 33,860
REFERENCE/DOCKET NUMBER: 48147/1699-CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-523-3400
TELEFAX: 617-523-6440
INFORMATION FOR SEQ ID NO: 29:
SEQUENCE CHARACTERISTICS:
LENGTH: 182 amino acids
TYPE: amino acid
STRANDEDNESS: linear
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-045-632-29

Query Match 14.0%; Score 97; DB 3; Length 182;
Best Local Similarity 28.1%; Pred. No. 0.00028;
Matches 25; Conservative 21; Mismatches 41; Indels 2; Gaps 2;

QY 21 TGNRYTTLAFLLEGAPMPGFTLKGGHEPLISVEEGKADTSSKLGADDEVVHIN 80
DB 94 SSGAIIYVELKRYGGPLGITISGTEPPDPIIISLTKGLAER-TGAIHIDRLAIN 152
QY 81 EYTLSSS-RKAVSLVKSYSKTLRLVVR 108
DB 153 SSSLKGRKPLSEDIHLLQMGFTVTLKIKK 181

Search completed: January 31, 2002, 20:57:10
Job time: 53 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 20:59:17 ; Search time 40.57 Seconds
(without alignments)
249.722 Million cell updates/sec

Title: US-09-641-831-6
Perfect score: 695
Sequence: 1 MMRTEDFHKPSATLNSNTA.....CCPEKVVAIHRDHDHALQP 133

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: PIR 68: *
2: PIR 1: *
3: PIR 2: *
4: PIR 4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	280	40.3	1616	I37183	gene APXL protein - human
2	131.5	18.9	913	A41109	protein-tyrosine-p
3	127.5	18.3	1171	T42372	probable guanylate
4	126	18.1	852	T10811	channel associated
5	126	18.1	870	G01974	channel associated
6	123.5	17.8	1256	JE0209	brain-specific ang
7	118.5	17.1	327	JC4385	IRM protein - rat
8	116.5	16.8	330	S71828	gene rll protein -
9	112.5	16.2	911	I56552	synapse-associated
10	112.5	16.2	926	I44105	protein-tyrosine-p
11	111	16.0	904	I38757	homolog of Drosoph
12	111	16.0	926	T27179	homolog of Drosoph
13	110.5	15.9	317	T27179	hypothetical prote
14	108.5	15.6	1277	T23160	synaptic scaffold
15	108	15.5	1012	T23160	hypothetical prote
16	107.5	15.5	455	A55050	enigma - human
17	105.5	15.2	517	T47134	hypothetical prote
18	104.5	15.0	720	A45436	synapse-associated
19	104.5	15.0	724	JH0800	postsynaptic densi
20	104.5	15.0	767	T09599	postsynaptic densi
21	103	14.8	624	T19630	hypothetical prote
22	103	14.8	1026	T19631	hypothetical prote
23	103	14.8	1112	T32733	AMPA glutamate rec
24	101.5	14.6	431	T16191	hypothetical prote
25	99	14.2	1367	T13703	tama protein - fru
26	98.5	14.2	423	T21570	hypothetical prote
27	98.5	14.2	440	T21568	hypothetical prote
28	98.5	14.2	960	A39651	discs-large tumor
29	98.5	14.2	1387	JC5502	G-protein signalin

30 97 14.0 505 2 S62894 alpha-syntrophin -
31 96 13.8 723 2 T14765 hypothetical prote
32 95 13.7 503 2 T84771 syntrophin-1 - mou
33 95 13.7 505 2 A53214 dystrophin-associa
34 94.5 13.6 390 2 T28036 hypothetical prote
35 94.5 13.6 1131 2 T15617 hypothetical prote
36 93 13.4 538 2 I59291 betai-syntrophin -
37 90.5 13.0 628 2 T09458 numb-binding prote
38 90.5 13.0 728 2 T09457 numb-binding prote
39 88.5 12.7 1495 2 T31434 desmin-180 - rat
40 85 12.2 2450 2 S71625 protein-tyrosine-p
41 84.5 12.2 488 2 I51379 syntrophin - Pacif
42 84 12.1 2294 2 I67630 protein tyrosine p
43 84 12.1 2466 2 I67629 protein tyrosine p
44 84 12.1 2490 1 A54971 protein-tyrosine-p
45 82.5 11.9 87 2 S60315 PSD-95-related pro

ALIGNMENTS

RESULT 1
I37183
gene APXL protein - human
C:Species: Homo sapiens (man)
C:Date: 29-May-1998 #sequence_revision 29-May-1998 #text_change 21-Jul-2000
C:Accession: I37183
R:Schiaffino, M.V.; Bassi, M.T.; Rugarli, E.I.; Renieri, A.; Galli, L.; Ballabio, A.
Hum. Mol. Genet. 4, 373-382, 1995
A:Title: Cloning of a human homologue of the Xenopus laevis APX gene from the ocular
A:Reference number: I37183; MUID:95315933
A:Accession: I37183
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: mRNA
A:Residues: 1-1616 <RES>
A:Cross-references: EMBL:X83543; NID:g790999; PIDN:CAA58534.1; PID:g1181628
C:Genetics:
A:Gene: GDB:APXL
A:Cross-references: GDB:582527
A:Map position: Xp22.3-Xp22.3
C:Superfamily: GLGF domain homology
F:32-104/Domain: GLGF domain homology <GLG>

Query Match 40.3%; Score 280; DB 2; Length 1616;
Best Local Similarity 52.5%; Pred. No. 1.le-19;
Matches 62; Conservative 15; Mismatches 29; Indels 12; Gaps 4;
QY 13 ATLSNNTATKGRVYILEAFLEGGAPWGFTLKGLGHEGLIISKVEGGKADTLSSKLQA 72
Db 14 AEATRAADGRLV--EVQLSGGAPWGFTLKGGEGHEGLVITKIEGSKAAAV-DKLLA 70
QY 73 GDEVVHINEVTLSSSRKEAVSLKGSYKTLRLVVRNGVLLCCPEKVVAIHRDHA 130
Db 71 GDEIVGINDIGLSGFRQEAICLVKGSHTKLVLVVRSEL-----GWR-----PHSWHA 119

RESULT 2
A41109
protein-tyrosine-phosphatase (EC 3.1.3.48), nonreceptor type 3 - human
N:Alternate names: PTPH1
C:Species: Homo sapiens (man)
C:Date: 27-Mar-1992 #sequence_revision 02-May-1994 #text_change 11-Jun-1999
C:Accession: A41109; I55698
R:Yang, Q.; Tonks, N.K.
Proc. Natl. Acad. Sci. U.S.A. 88, 5949-5953, 1991
A:Title: Isolation of a cDNA clone encoding a human protein-tyrosine phosphatase with
A:Reference number: A41109; MUID:91296738
A:Accession: A41109
A:Molecule type: mRNA
A:Residues: 1-913 <YAN>
A:Cross-references: GB:M64572; NID:g179912; PIDN:AAA35647.1; PID:g179913
R:Ikuta, S.; Itoh, F.; Hinoda, Y.; Toyota, M.; Makiguchi, Y.; Imai, K.; Yachi, A.

J. Gastroenterol. 29, 727-732, 1994
 A:Title: Expression of cytoskeletal-associated protein tyrosine phosphatase PRPH1 mRNA
 A:Reference number: 155698; MUID:95179278
 A:Accession: 155698
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 899-913 <RES>
 A:Cross-references: GB:S76309; NID:9913165; PIDN:AA33583.1; PID:9913166
 C:Genetics:
 A:Gene: GDB:PTPN3
 A:Cross-references: GDB:131386; OMIM:176877
 A:Map position: 9q31-9q31
 C:Superfamily: protein-tyrosine-phosphatase, nonreceptor type 3; GLGF domain homology; F
 C:Keywords: phosphoprotein; phosphoric monoester hydrolase; tyrosine-specific phosphatase
 F:31-308/Domain: protein 4.1 membrane-binding domain homology <B41>
 F:516-590/Domain: GLGF domain homology <GLG>
 F:670-890/Domain: protein-tyrosine-phosphatase homology <PTP>
 F:842/Active site: Cys (phosphocysteine intermediate) #status predicted
 F:848/Binding site: substrate phosphate (Arg) #status predicted

Query Match 18.9%; Score 131.5; DB 1; Length 913;
 Best Local Similarity 28.5%; Pred. No. 4, 8e-05;
 Matches 35; Conservative 23; Mismatches 46; Indels 19; Gaps 4;
 QY 6 EDHKKPATLNSMTATKGR-----IYLEAFLEGAPMGFTLKGLHEGPELI 53
 DB 479 DDFRR--VTGGSTEDASQYYCDKNDGSLYLIRITTPEDCKFGNKGVDQKMPLV 536
 QY 54 ISKVEEGKADTLSSKLAQDEVDVINEVTLS-SRKEAVSLVKGSKYT---LRLVVR 108
 DB 537 VSRINPSPADTCIPKINBDQVILINGRDISHTHQVVMFKASHSRRLALVIR 596
 QY 109 NGV 111
 DB 597 RAV 599

RESULT 3
 T42372
 probable guanylate kinase (EC 2.7.4.8) 1, membrane-associated, splice form b - mouse
 C:Species: Mus musculus (house mouse)
 C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 18-Aug-2000
 C:Accession: T42372
 R:DoProctokaya, I.; Guy, R. K.; James, G. L.
 J. Biol. Chem. 272, 31589-31597, 1997
 A:Title: Mg1-1: a membrane-associated guanylate kinase with a unique arrangement of pro
 A:Reference number: 222139; MUID:98058950
 A:Accession: T42372
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 11171 <DOB>
 A:Cross-references: EMBL:AF027503; NID:q2702346; PID:q2702347; PIDN:AA91995.1
 A:Experimental source: strain C57 Black/6 x CBA
 C:Genetics:
 A:Gene: Mag1-1
 C:Superfamily: WW repeat homology
 C:Keywords: alternative splicing; phosphotransferase
 F:300-337/Domain: WW repeat homology <KWR1>
 F:347-384/Domain: WW repeat homology <KWR2>

Query Match 18.3%; Score 127.5; DB 2; Length 1171;
 Best Local Similarity 27.3%; Pred. No. 0.00016;
 Matches 39; Conservative 25; Mismatches 52; Indels 27; Gaps 6;
 QY 9 HKPS--ATLNSMTATKGR-----IYLEAFLEGAPMGFTLKGLHEGE 50
 DB 998 HAPSOOGQJOETRTTKPKQDOSFEKFGQAQOEDFYVELERGAKGFGFSLRGREYNM 1057
 QY 51 PLITKVEEGKADTLSSKLAQDEVDVINEVTLS-SRKEAVSLVKGSKYTTLRLVVRN 109
 DB 1058 DLYVRLAEDGPAER-CGMRIGDEILINGETTKNMKHSRAIELIKNGSRVRLPLRG 1116

QY 110 GVLCCPEW-----KVVAIRHD 127
 DB 1117 DGSV--PEYAMIPPKIAMCRNE 1137

RESULT 4
 T10811
 channel associated protein of synapse 2 - rat
 C:Species: Rattus norvegicus (Norway rat)
 C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jan-2000
 C:Accession: T10811
 R:Irle, M.; Hata, Y.; Takai, Y.
 submitted to the EMBL Data Library, April 1996
 A:Description: Cloning of new isoforms of PSD-95/SAP90 related genes.
 A:Reference number: 217166
 A:Accession: T10811
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-852 <IRI>
 A:Cross-references: EMBL:U53368; NID:q1517939; PID:q1517940
 C:Genetics:
 A:Note: PSD-95/SAP90-related gene 2 (chapsyn isoform2)
 C:Superfamily: discs-large tumor suppressor; GLGF domain homology; guanylate kinase h
 F:198-276/Domain: GLGF domain homology <GLG>
 F:543-601/Domain: SH3 homology <SH3>
 F:663-840/Domain: guanylate kinase homology <GKI>

Query Match 18.1%; Score 126; DB 2; Length 852;
 Best Local Similarity 32.7%; Pred. No. 0.00016;
 Matches 33; Conservative 26; Mismatches 34; Indels 8; Gaps 5;
 QY 10 KPSATLNSMTATK--RYILEAFLEGAPMGFTLKGLHEGPELIISKVEEGKADTL 67
 DB 404 QPSVTLQRAISLEGEPRKVLH---KSTGLGNINIVGG-EDSGIVSFILAGPAD-LS 458
 QY 68 SKLAQDEVDVINEVTLS-SRKEAVSLVKGSKYTTLRLVVR 107
 DB 459 GEIQRGQILSVNGIDLRGASHQAAALAKGACQTVIIMQ 499

RESULT 5
 G01974
 channel associated protein of synapse - human
 C:Species: Homo sapiens (man)
 C:Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 21-Jan-2000
 C:Accession: G01974
 R:Kim, E.; Cho, K.
 submitted to the EMBL Data Library, July 1995
 A:Reference number: G08966
 A:Accession: G01974
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-870 <KIM>
 A:Cross-references: EMBL:U32376; NID:q1463025; PID:q1036790
 C:Genetics:
 A:Gene: chapsyn-110
 C:Superfamily: discs-large tumor suppressor; GLGF domain homology; guanylate kinase h
 F:198-276/Domain: GLGF domain homology <GLG2>
 F:543-601/Domain: SH3 homology <SH3>
 F:681-858/Domain: guanylate kinase homology <GKI>

Query Match 18.1%; Score 126; DB 2; Length 870;
 Best Local Similarity 32.7%; Pred. No. 0.00016;
 Matches 33; Conservative 26; Mismatches 34; Indels 8; Gaps 5;
 QY 10 KPSATLNSMTATK--RYILEAFLEGAPMGFTLKGLHEGPELIISKVEEGKADTL 67
 DB 404 QPSVTLQRAVSLBEGEPKRVLH---KSTGLGNINIVGG-EDSGIVSFILAGPAD-LS 458
 QY 68 SKLAQDEVDVINEVTLS-SRKEAVSLVKGSKYTTLRLVVR 107

Query Match	17.1%;	Score 118.5;	DB 2;	Length 327;
Best Local Similarity	39.2%;	Pred. No. 0.00028;		
Matches	31;	Conservative	12;	Mismatches 33;
				Indels 3;
				Gaps 2;

Query Match	16.2%;	Score 112.5;	DB 2;	Length 911;
Best Local Similarity	27.9%;	Pred. No. 0.0039;		
Matches 34;	Conservative 27;	Mismatches 42;	Indels 19;	Gaps 5;

Qy 3 RTTEDFHKPSATLNSNTATKGRYYL-EAFL-----EGGAPWGFTLKGGL 46

[illegible]

```

F:324-402/Domain: GLGF domain homology <GLGZ>
F:588-646/Domain: SH3 homology <SH3>
F:715-892/Domain: guanylate kinase homology <GKI>

Query Match          16.0%; Score 111; DB 2; Length 904;
Best Local Similarity 30.3%; Pred. No. 0.0055;
Matches 36; Conservative 23; Mismatches 40; Indels 20; Gaps 6;

OY      7 DPH-KPSATLNSNTATKGRIVYL-EAFV-----EGGAPWGTLLKGLEHG 49
        | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      428 DNHWSPSSFLGOTPPASPARISPYSKAVLGDDDETRPRKVVLHRSSTGLGNIVGG-EDG 486
        | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY      50 EPLIISKVEEGKADPTLSKLQAGDEVHINEVTL-SSSRKEAVSLVGSKYTLLRVLR 107
        | : | : | : | | | | | : | : | : | : | : | : | : | : | : |
Db      487 EGIFISFIILAGPAD-LSGELRKGDRIITSVNSVDLRAASHDQAALAKNAQANTIVAQ 544
        | : | : | : | | | | | : | : | : | : | : | : | : | : | : |

RESULT 12
homolog of Drosophila discs large protein, isoform 2 - human
138756
C:Species: Homo sapiens (man)
C>Date: 01-Mar-1996 #sequence_revision 01-Mar-1996 #text_change 21-Jan-2000
C:Accession: I38756
R:Luc, R.A.; Marfatia, S.M.; Branton, D.; Chishti, A.H.
Proc. Natl. Acad. Sci. U.S.A. 91, 9818-9822, 1994
A>Title: Cloning and characterization of hdlg: the human homologue of the Drosophila
A:Reference number: I38756; MUID:95024052
A:Accession: I38756
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-926 <RES>
A:Cross-references: EMBL:U13896; NID:g558435; PIDD:AAA50598.1; PID:g558436
C:Genetics:
A:Gene: GDB:DLG1
A:Cross-references: GDB:393278; OMIM:601014
A:Map position: 3q29-3q29
C:Superfamily: discs-large tumor suppressor; GLGF domain homology; guanylate kinase h
C:Keywords: alternative splicing; duplication
F:229-307/Domain: GLGF domain homology <GLGI>
F:324-402/Domain: GLGF domain homology <GLGZ>
F:588-646/Domain: SH3 homology <SH3>
F:737-914/Domain: guanylate kinase homology <GKI>

Query Match          16.0%; Score 111; DB 2; Length 926;
Best Local Similarity 30.3%; Pred. No. 0.0057;
Matches 36; Conservative 23; Mismatches 40; Indels 20; Gaps 6;

OY      7 DPH-KPSATLNSNTATKGRIVYL-EAFV-----EGGAPWGTLLKGLEHG 49
        | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      428 DNHWSPSSFLGOTPPASPARISPYSKAVLGDDDETRPRKVVLHRSSTGLGNIVGG-EDG 486
        | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY      50 EPLIISKVEEGKADPTLSKLQAGDEVHINEVTL-SSSRKEAVSLVGSKYTLLRVLR 107
        | : | : | : | | | | | : | : | : | : | : | : | : | : | : |
Db      487 EGIFISFIILAGPAD-LSGELRKGDRIITSVNSVDLRAASHDQAALAKNAQANTIVAQ 544
        | : | : | : | | | | | : | : | : | : | : | : | : | : | : |

RESULT 13
T27179
hypothetical protein Y54G11A.10 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T27179
R:Wallis, J.
submitted to the EMBL Data Library, December 1998
A:Reference number: Z20322
A:Accession: T27179
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-317 <MID>
A:Cross-references: EMBL:AL034488; NID:e1359895; PIDD:CMA22459.1; CESP:Y54G11A.10
A:Experimental source: clone Y54G11A
```

C:Genetics:
A:Gene: CESP:Y54G11A.10
A:Introns: 55/3; 108/3; 175/3; 228/2; 253/3
Query Match 15.9%; Score 110.5; DB 2; Length 317;
Best Local Similarity 31.0%; Pred. No. 0.0017; Indels 9; Gaps 4;
Matches 31; Conservative 24; Mismatches 36; Indels 9; Gaps 4;
A:Introns: 107/1; 151/2; 183/2; 270/2; 316/3; 410/3; 628/3; 922/3; 971/3
C:Superfamily: WW repeat homology
F:131-168/Domain: WW repeat homology <WWR>
Query Match 15.5%; Score 108; DB 2; Length 1012;
Best Local Similarity 29.1%; Pred. No. 0.013; Indels 20; Gaps 5;
Matches 34; Conservative 24; Mismatches 39; Indels 20; Gaps 5;
QY 11 PSATLNSNTATKG----RYIVLEAF-----LEGGAPMGFTLKGGLHGEPLII 54
DB 644 PASRMRPSTTLTGFPATPNYIPLSQYNQKPSDLITVSLIRKPVGFGFRLGGVESKTPLSV 703
QY 55 SKVEEGKADTLSSKLOAGDEVVHINEVTLS--SSSRKEAVSLVKGS--YKTLRLVVR 108
DB 704 GOIVIGGAAEE--DGRLOEGDEIVEIDGHNVEGASHSEAVVLEAAQNKHKVLIYVR 759
Search completed: January 31, 2002, 20:59:17
Job time: 65 sec

C:Genetics:
A:Gene: CESP:Y54G11A.10
A:Introns: 55/3; 108/3; 175/3; 228/2; 253/3
Query Match 15.9%; Score 110.5; DB 2; Length 317;
Best Local Similarity 31.0%; Pred. No. 0.0017; Indels 9; Gaps 4;
Matches 31; Conservative 24; Mismatches 36; Indels 9; Gaps 4;
A:Introns: 107/1; 151/2; 183/2; 270/2; 316/3; 410/3; 628/3; 922/3; 971/3
C:Superfamily: WW repeat homology
F:131-168/Domain: WW repeat homology <WWR>
QY 13 ATLNSNTATKG----RYIVLEAFLEGGAPMGFTLKGGLHGEPLIIISKVEEGKADTLSS 68
DB 184 ATVAFAAEGHAPRIVELPKTDQ--LGFNVMGKQKQNSPIIYSRIIPGVADR-HG 239
QY 69 KLOAGDEVVHINEVTLS--RKEAVSLVSGSKYTLRLVVR 107
DB 240 GLKRGDOLIAVNGVNAEACHEKAVDLLKSAVGSVKLVIR 279
RESULT 14
T14152
synaptic scaffolding protein S-SCAM - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 18-Aug-2000
C:Accession: T14152
R:Hirao, K.; Hata, Y.; Ide, N.; Takeuchi, M.; Irie, M.; Yao, I.; Deguchi, M.; Toyoda, A.
J. Biol. Chem. 273, 21105-21110, 1998
A:Title: A novel multiple PDZ domain-containing molecule interacting with N-methyl-D-aspartate receptor 1
A:Reference number: Z17889; MUID:98361985
A:Accession: T14152
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1277 <HR>
A:Cross-references: EMBL:AF034863; NID:93411052; PID:93411053; PIDN:RAC31124.1
C:Genetics:
A:Gene: S-SCAM
C:Function:
A:Description: may assemble receptors and cell adhesion proteins at synaptic junctions
C:Superfamily: WW repeat homology
F:302-339/Domain: WW repeat homology <WW1>
F:348-385/Domain: WW repeat homology <WW2>
Query Match 15.6%; Score 108.5; DB 2; Length 1277;
Best Local Similarity 29.8%; Pred. No. 0.015; Indels 3; Gaps 3;
Matches 25; Conservative 24; Mismatches 32; Indels 3; Gaps 3;
QY 27 YLEAFLEGGAP-WGFTLKGGLHGEPLIIISKVEEGKADTLSSKLOAGDEVVHIN-EVTL 84
DB 1139 YFTVDMKGAAGFGFSGRGREYKMDLYVRLAEDGPA-IRNGRMVGDQIIIEINGESTR 1197
QY 85 SSSRKEAVSLVKGSYKTLRLVVR 108
DB 1198 DMTHARIELIKSGGRRVRLLLKR 1221
RESULT 15
T23160
hypothetical protein K01A6.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Aug-2000
C:Accession: T23160
R:Cottage, A.
submitted to the EMBL Data Library, January 1996
A:Reference number: Z19701
A:Accession: T23160
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1012 <WIL>
A:Cross-references: EMBL:Z68750; PIDN:CAA92965.1; GSPDB:GN00022; CESP:K01A6.1
A:Experimental source: clone K01A6
C:Genetics:
A:Gene: CESP:K01A6.1
A:Map position: 4

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 21:06:28 ; Search time 26.99 Seconds

(without alignments)
180.675 Million cell updates/sec

Title: US-09-641-831-6

Perfect score: 695

Sequence: 1 MMRTEDFHKPSATLSNTA.....CCPEKVVAIHRDHALQP 133

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query No.	Score	Match	Length	DB	ID	Description
1	280	40.3	1616	1	APXL_HUMAN	Q13796	homo sapien
2	131.5	18.9	913	1	PTN3_HUMAN	P26045	homo sapien
3	126	18.1	852	1	DLG2_RAT	Q63622	rattus norv
4	126	18.1	870	1	DLG2_HUMAN	P15700	homo sapien
5	118.5	17.1	327	1	PDLL_RAT	P52944	rattus norv
6	116.5	16.8	330	1	RIL_RAT	P36202	rattus norv
7	112.5	16.2	330	1	RIL_CHICK	Q9PW72	gallus gall
8	112.5	16.2	911	1	DLG1_RAT	Q62696	rattus norv
9	112.5	16.2	926	1	PTN4_HUMAN	P29074	homo sapien
10	111.5	16.0	849	1	DLG3_MOUSE	P70175	mus musculus
11	111.5	16.0	849	1	DLG3_RAT	Q62936	rattus norv
12	111	16.0	904	1	DLG1_HUMAN	Q12959	homo sapien
13	109.5	15.8	329	1	PDLL_HUMAN	O00151	homo sapien
14	104.5	15.0	724	1	DLG4_MOUSE	Q62108	mus musculus
15	104.5	15.0	724	1	DLG4_RAT	P31016	rattus norv
16	104.5	15.0	767	1	DLG4_HUMAN	P78352	homo sapien
17	104	15.0	817	1	DLG3_HUMAN	Q92796	homo sapien
18	103	14.8	1026	1	PTPL_CAEEL	P28191	caenorhabdi
19	101	14.5	326	1	PDLL_MOUSE	O70400	mus musculus
20	99	14.2	328	1	RIL_HUMAN	P50479	homo sapien
21	98.5	14.2	960	1	DLG1_DROME	P31007	drosophila
22	98.5	14.2	1387	1	RGSC_RAT	O08774	rattus norv
23	93	13.4	933	1	ZO3_HUMAN	Q95049	homo sapien
24	92.5	13.3	1447	1	RGSC_HUMAN	O14924	homo sapien
25	88.5	12.7	898	1	ZO3_CANFA	O62683	canis famli
26	84	12.1	2485	1	PTND_HUMAN	Q12923	homo sapien
27	83	11.9	643	1	RHOP_MOUSE	Q61085	mus musculus
28	81	11.7	1167	1	ZO2_MOUSE	Q920u1	mus musculus
29	80.5	11.6	631	1	IL15_HUMAN	Q14005	homo sapien
30	80	11.5	585	1	MPP3_HUMAN	Q13368	homo sapien
31	79.5	11.4	905	1	ZO3_MOUSE	Q9qxy1	mus musculus
32	77	11.1	1190	1	ZO2_HUMAN	Q9udy2	homo sapien
33	77	11.1	1736	1	ZO1_HUMAN	Q07157	homo sapien

34	77	11.1	1745	1	ZO1_MOUSE	P39447	mus musculus
35	76	10.9	1174	1	ZO2_CANFA	Q95168	canis famli
36	75.5	10.9	206	1	OM25_RAT	O9WV14	rattus norv
37	75	10.8	3988	1	POLG_BVDVN	P19711	bovine vira
38	74.5	10.7	837	1	APBL_RAT	Q02410	homo sapien
39	74.5	10.7	839	1	APBL_RAT	O35430	rattus norv
40	74	10.6	3898	1	POLG_BVDVS	Q01499	bovine vira
41	72.5	10.4	145	1	OM25_HUMAN	P57105	homo sapien
42	72.5	10.4	621	1	TRA_STRLI	P22409	streptomyce
43	72	10.4	322	1	CYS1_HOMAM	P13277	homo sapien
44	72	10.4	397	1	DJAI_HUMAN	P31689	homo sapien
45	72	10.4	397	1	DJAI_MOUSE	P54102	mus musculus

ALIGNMENTS

RESULT 1	APXL_HUMAN	STANDARD;	PRT; 1616 AA.
ID	APXL_HUMAN	STANDARD;	PRT; 1616 AA.
AC	Q13796;		
DT	01-NOV-1997 (Rel. 35, Created)		
DT	01-NOV-1997 (Rel. 35, Last sequence update)		
DT	20-AUG-2001 (Rel. 40, Last annotation update)		
DE	APICAL-LIKE PROTEIN (APXL PROTEIN).		
GN	APXL.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=Retina;		
RA	Schiaffino V.M., Bassi M.T., Rugarli E.I., Renieri A., Galli L.,		
RA	Ballabio A.;		
RT	"Cloning of a human homologue of the Xenopus laevis APX gene from the		
RT	ocular albinism type 1 critical region.";		
RL	Hum. Mol. Genet. 4:373-382(1995).		
RN	[2]		
RP	SEQUENCE OF 56-1616 FROM N.A.		
RA	Shen Y., Gibbs R.A.;		
RL	Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.		
CC	-!- TISSUE SPECIFICITY: ABUNDANT IN RETINA AND MELANOMA; ALSO IN		
CC	BRAIN, PLACENTA, LUNG, KIDNEY, PANCREAS.		
CC	-!- SIMILARITY: SOME, TO XENOPUS LAEVIS APICAL PROTEIN (APX).		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/		
CC	or send an email to license@isb-sib.ch).		
CC	-----		
DR	EMBL; X83543; CAA58534.1;		
DR	EMBL; AC002365; AAC32592.1;		
DR	HSSP; Q12959; 1PDR.		
DR	MIM; 300103;		
DR	InterPro; IPR000857; IGFBP.		
DR	InterPro; IPR001478; PDZ.		
DR	Pfam; PF00219; IGFBP; 1.		
DR	Pfam; PF00595; PDZ; 1.		
DR	SMART; SM00228; PDZ; 1.		
DR	PROSITE; PS50106; PDZ; 1.		
DR	SIMILAR 54 112		
FT	DOMAIN 150 153		TO MOUSE SYNTROPHIN-1 AND RABBIT
FT	DOMAIN 314 320		DYSTROPHIN-ASSOCIATED PROTEIN.
FT	DOMAIN 343 346		POLY-PRO.
FT	DOMAIN 1065 1068		POLY-ALA.
FT	DOMAIN 1065 1068		POLY-PRO.
FT	SEQUENCE 1616 AA; 176409 MW; 752406B5BC0B60A2 CRC64;		

Query Match 40.3%; Score 280; DB 1; Length 1616;
 Best Local Similarity 52.5%; Pred. No. 9,1e-20;
 Matches 62; Conservative 15; Mismatches 29; Indels 12; Gaps 4;

DB 13 ALNSTATKGRVYIEAFLEGAPMGFTLKGLEGEPLISKEVEEGKADTLSSKIQ 72
 14 AEAEETAAAGGRLV--EVLQSGAPMGFTLKGREHGEPLVITKEEGSKAAAV-DKLLA 70
 73 GDEYVHINEVTLSSRKREKSVLKSGYKTLVVRNGVLLCCPEKKVVAIRRHQHA 130
 71 GDIVGIVINDIGLSGFEQEAICLVKSGSHKTLKLVKRSEL---GWR-----PSHMA 119

RESULT 2
 PTN3_HUMAN STANDARD; PRT; 913 AA.
 ID PTN3_HUMAN
 AC P26045;
 DT 01-MAY-1992 (Rel. 22, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE PROTEIN-TYROSINE PHOSPHATASE, NON-RECEPTOR TYPE 3 (EC 3.1.3.48)
 DE (PROTEIN-TYROSINE PHOSPHATASE H) (PTP-H1).
 GN PTPN3 OR PTPH1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 OC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91296738; PubMed=1648725;
 RA Yang Q., Tooks N.K.;
 RT Isolation of a cDNA clone encoding a human protein-tyrosine
 RT phosphatase with homology to the cytoskeletal-associated proteins
 RT band 4.1, ezrin, and talin.";
 RL Proc. Natl. Acad. Sci. U.S.A. 88:5949-5953(1991).
 RN [2]
 RP SEQUENCE OF 194-896 FROM N.A.
 RC TISSUE-Colon;
 RX MEDLINE=9327504; PubMed=1626183;
 RA Airmura Y., Hinoda Y., Itoh F., Takekawa M., Tsujisaki M., Adachi M.,
 RA Imai K., Yachi A.;
 RT "cDNA cloning of new protein tyrosine phosphatases in the human
 RT colon.";
 RL Tumour Biol. 13:180-186(1992).
 RN [3]
 RP SEQUENCE OF 899-913 FROM N.A.
 RX MEDLINE=95179278; PubMed=7874267;
 RA Ikuta S., Itoh F., Hinoda Y., Toyota M., Makiguchi Y., Imai K.,
 RA Yachi A.;
 RT "Expression of cytoskeletal-associated protein tyrosine phosphatase
 RT PTPH1 mRNA in human hepatocellular carcinoma.";
 RL J. Gastroenterol. 29:727-732(1994).
 CC -1- FUNCTION: MAY ACT AT JUNCTIONS BETWEEN THE MEMBRANE AND THE
 CC CYTOSKELETON.
 CC -1- CATALYTIC ACTIVITY: PROTEIN-TYROSINE PHOSPHATE + H(2)O =
 CC PROTEIN-TYROSINE + ORTHOPHOSPHATE.
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
 CC -1- SIMILARITY: BELONGS TO THE NON-RECEPTOR CLASS OF THE PROTEIN-
 CC TYROSINE PHOSPHATASE FAMILY.
 CC -1- SIMILARITY: CONTAINS 1 PDZ/4.1-LIKE DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M64572; AAA35647.1; -
 CC EMBL: S39392; AAB22439.2; -
 CC DR

EMBL: S76309; AAB33583.1; -
 DR PIR: A41109; A41109.
 DR HSSP: P18031; 1PTT.
 DR MIM: 176877; -
 DR InterPro: IPR000299; Band_4.1.
 DR InterPro: IPR001478; PDZ.
 DR InterPro: IPR000387; Tyr_phosphatase.
 DR InterPro: IPR000242; Tyr_prot_phphatase.
 DR Pfam: PF00373; Band_41; 1.
 DR Pfam: PF00595; PDZ; 1.
 DR Pfam: PF00102; Y_phosphatase; 1.
 DR PRINTS: PR00700; PTYPHPTASE.
 DR PRINTS: PR00935; BANDA1.
 DR SMART: SM00295; B41; 1.
 DR SMART: SM00228; PDZ; 1.
 DR SMART: SM00194; PTPe; 1.
 DR PROSITE: PS00660; BAND_41_1; 1.
 DR PROSITE: PS00661; BAND_41_2; 1.
 DR PROSITE: PS50057; BAND_41_3; 1.
 DR PROSITE: PS50106; PDZ; 1.
 DR PROSITE: PS00383; TYR_PHOSPHATASE_1; 1.
 DR PROSITE: PS50056; TYR_PHOSPHATASE_2; 1.
 DR PROSITE: PS50055; TYR_PHOSPHATASE_PTP; 1.
 DR PROSITE: PS50055; TYR_PHOSPHATASE_PTP; 1.
 KW Structural protein; Cytoskeleton; Hydrolase.
 FT DOMAIN 84 241
 FT DOMAIN 510 582
 FT DOMAIN 670 913
 FT ACT_SITE 842 842
 FT ACT_SITE 842 842
 FT BY_SIMILARITY.
 SQ SEQUENCE 913 AA; 104029 MW; 29A539ACDE2P1515 CRC64;

Query Match 18.9%; Score 131.5; DB 1; Length 913;
 Best Local Similarity 28.5%; Pred. No. 2.8e-05;
 Matches 35; Conservative 23; Mismatches 46; Indels 19; Gaps 4;

DB 6 EDFHKPSATLNSMTATKGR-----TYLEAFLEGAPMGFTLKGLEGEPLI 53
 479 DDFHR--VTGSGSTEDASOYYCDKNDGSLYLITPTDEGKREFNLKSGVDKMLV 536
 54 ISKVEEGKADTLSSKIQAGDEVHINEVTLSS-SRKEAVSLVKSQYKTLVVR 108
 537 VSRINPESPADTCTPRKLNEDQIVLNGRDISHHTDOYVMFTKASRSHRELALVIR 596

DB 109 NGV 111
 597 RAV 599

RESULT 3
 DLG2_RAT STANDARD; PRT; 852 AA.
 ID DLG2_RAT
 AC Q63622; Q62939; P70548;
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE CHANNEL ASSOCIATED PROTEIN OF SYNAPSE-110 (CHAPSYN-110) (SYNAPTIC
 DE DENSITY PROTEIN PSD-93) (DISCS, LARGE HOMOLOG 2).
 GN DLG2.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OC NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96310881; PubMed=8755482;
 RA Kim E., Cho K.O., Rothschild A., Sheng M.,
 RA "Heteromultimerization and NMDA receptor-clustering activity of
 RA Chapsyn-110, a member of the PSD-95 family of proteins.";
 RL Neuron 17:103-113(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96193770; PubMed=8625413;
 RA Brenman J.E., Chao D.S., Gee S.H., McGee A.W., Craven S.E.,

RA Santillano D.R., Wu Z., Huang F., Xia H., Peters M.F.,
 RA Froehner S.C., Bredt D.S.;
 RT Interaction of nitric oxide synthase with the postsynaptic density
 RT protein PSD-95 and alpha1-syntrophin mediated by PDZ domains.";
 RL Cell 84:757-767(1996).
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Irie M., Hata Y., Takai Y.;
 RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: INTERACTS WITH THE CYTOPLASMIC TAIL OF NMDA RECEPTOR
 CC SUBUNITS AS WELL AS POTASSIUM CHANNELS.
 CC -1- SIMILARITY: CONTAINS 3 PDZ/DHR DOMAINS.
 CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
 CC -1- SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U49049; AAB53243.1; -
 CC EMBL: U50717; AAC52643.1; -
 CC EMBL: U53368; AAB48562.1; -
 DR HSSP: Q12959; 1PDR.
 DR InterPro: IPR000619; Guanylate_kin.
 DR InterPro: IPR001478; PDZ.
 DR InterPro: IPR001452; SH3.
 DR Pfam: PF00625; Guanylate_kin; 1.
 DR Pfam: PF00595; PDZ; 3.
 DR Pfam: PF00018; SH3; 1.
 DR SMART: SM00072; Gukc; 1.
 DR SMART: SM00228; PDZ; 3.
 DR SMART: SM00326; SH3; 1.
 DR PROSITE: PS00856; GUANYLATE_KINASE_1; 1.
 DR PROSITE: PS50052; GUANYLATE_KINASE_2; 1.
 DR PROSITE: PS50106; PDZ; 3.
 DR PROSITE: PS50002; SH3; 1.
 DR SH3 domain; Repeat.
 FT DOMAIN 98 184 PDZ 1.
 FT 193 279 PDZ 2.
 FT DOMAIN 421 501 PDZ 3.
 FT 536 606 SH3.
 FT DOMAIN 662 852
 FT 181 182 GUANYLATE_KINASE.
 FT CONFLICT 228 228 VR -> IL (IN REF. 2).
 FT CONFLICT 326 326 I -> M (IN REF. 2).
 FT CONFLICT 339 339 R -> K (IN REF. 2).
 FT CONFLICT 450 450 D -> E (IN REF. 3).
 FT CONFLICT 464 464 MISSING (IN REF. 2).
 FT CONFLICT 474 474 GD -> RK (IN REF. 2).
 FT CONFLICT 476 476 D -> H (IN REF. 2).
 FT CONFLICT 478 478 R -> P (IN REF. 2).
 FT CONFLICT 484 486 A -> D (IN REF. 2).
 FT CONFLICT 506 506 AAA -> LP (IN REF. 2).
 FT CONFLICT 569 569 A -> S (IN REF. 2).
 FT CONFLICT 586 586 H -> N (IN REF. 2).
 FT CONFLICT 626 641 L -> Q (IN REF. 2).
 FT CONFLICT 639 639 MISSING (IN REF. 2).
 FT CONFLICT 726 726 K -> A (IN REF. 3).
 FT CONFLICT 733 733 F -> L (IN REF. 1).
 FT CONFLICT 749 749 N -> Y (IN REF. 2).
 FT CONFLICT 756 756 E -> V (IN REF. 1).
 FT CONFLICT 791 792 L -> H (IN REF. 2).
 FT CONFLICT 794 794 KR -> NG (IN REF. 2).
 FT CONFLICT 794 794 T -> M (IN REF. 2).
 SQ SEQUENCE 852 AA; 94934 MM; F8D414A8B9C5B09 CRC64;

Matches 33; Conservative 26; Mismatches 34; Indels 8; Gaps 5;
 QY 10 KPSATLNSNTATKG--RYVLEAFLEGAPMGFTLKGLEGEPLIISKVEGGKADTL 67
 DB 404 OPSVTLQALISLEGPRKRVLH---KSGTGLGFNVGS-EDGEGLFVSEFLIAGGPAD-LS 458
 QY 68 SKLQAGDEVHINEVTL-SSSRKEAVSLVKGSKYKTLRYLR 107
 DB 459 GELRGDQILSYNGIDILRGASHEDQAAALKGAGQVTLIAQ 499
 RESULT 4
 ID DLG2_HUMAN STANDARD; PRT; 870 AA.
 AC Q15700;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DE 20-AUG-2001 (Rel. 40, Last annotation update)
 DE CHANNEL ASSOCIATED PROTEIN OF SYNAPSE-110 (CHAPSYN-110) (DISCS, LARGE
 DE HOMOLOG 2).
 GN DLG2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 OX NCBI_TaxId=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-96310881; PubMed-8755482;
 RA Kim E., Cho K.-O., Rothschild A., Sheng M.;
 RT "Heteromultimerization and NMDA receptor-clustering activity of
 RT Chapsyn-110, a member of the PSD-95 family of proteins.";
 RL Neuron 17:103-113(1996).
 CC -1- FUNCTION: INTERACTS WITH THE CYTOPLASMIC TAIL OF NMDA RECEPTOR
 CC SUBUNITS AS WELL AS POTASSIUM CHANNELS.
 CC -1- SIMILARITY: CONTAINS 3 PDZ/DHR DOMAINS.
 CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
 CC -1- SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U32376; AAB04949.1; -
 CC HSSP: Q12959; 1PDR.
 DR MIM: G03583; -
 DR InterPro: IPR000619; Guanylate_kin.
 DR InterPro: IPR001478; PDZ.
 DR InterPro: IPR001452; SH3.
 DR Pfam: PF00625; Guanylate_kin; 1.
 DR Pfam: PF00595; PDZ; 3.
 DR Pfam: PF00018; SH3; 1.
 DR SMART: SM00072; Gukc; 1.
 DR SMART: SM00228; PDZ; 3.
 DR SMART: SM00326; SH3; 1.
 DR PROSITE: PS00856; GUANYLATE_KINASE_1; 1.
 DR PROSITE: PS50052; GUANYLATE_KINASE_2; 1.
 DR PROSITE: PS50106; PDZ; 3.
 DR PROSITE: PS50002; SH3; 1.
 DR SH3 domain; Repeat.
 FT DOMAIN 98 184 PDZ 1.
 FT 193 279 PDZ 2.
 FT DOMAIN 421 501 PDZ 3.
 FT 536 606 SH3.
 FT DOMAIN 680 870 GUANYLATE_KINASE.
 SQ SEQUENCE 870 AA; 97500 MM; 89C83BA0619F6F59 CRC64;

Query Match 18.1%; Score 126; DB 1; Length 852;
 Best Local Similarity 32.7%; Pred. No. 9e-05;

Query Match 18.1%; Score 126; DB 1; Length 870;

Best Local Similarity 32.78; Pred. No. 9.2e-05;
Matches 33; Conservative 26; Mismatches 34; Indels 8; Gaps 5;
QY 10 KPSATLNNTATKG--RYIYLEAFLEGAPWGFTLKGLEHGEPLIISKVEGKADTLS 67
DB 404 QFSMTLQRAVSLGEPRKVLH--KGSTGLGFLNVGG-EDGEGIFVSLAGPAD-LS 458
QY 68 SKLQAGDEVHINVTLS--SSRKEAVSLVKSGLRLLVRR 107
DB 459 GELQRGDQILSYNGIDLRGASHEQAAALKGAGGTVTIIAQ 499

RESULT 5
RIL_RAT ID PDLL_RAT STANDARD; PRT; 327 AA.
AC P52944;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PDZ AND LIM DOMAIN PROTEIN 1 (LIM DOMAIN PROTEIN CLP-36) (C-TERMINAL
DE LIM DOMAIN PROTEIN 1) (ELFIN).
GN PDLLIM1 OR CLIM1 OR CLP36.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=Liver;
RX MEDLINE=96095533; PubMed=8522188;
RA Wang H., Harrison-Shostak D.C., Lemasters J.J., Herman B.;
RT "Cloning of a rat cDNA encoding a novel LIM domain protein with high
RT homology to rat RIL";
RL Gene 165:267-271(1995).
CC -1- FUNCTION: CYTOSKELETAL PROTEIN THAT MAY ACT AS AN ADAPTER THAT
CC BRINGS OTHER PROTEINS TO THE CYTOSKELETON (BY SIMILARITY).
CC -1- SUBUNIT: INTERACTS WITH ALPHA-ACTININS, 1 2 AND 4 (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC -1- TISSUE SPECIFICITY: EXPRESSED MOST ABUNDANTLY IN HEART, LUNG AND
CC LIVER, MODERATELY IN SPLEEN AND SKELETAL MUSCLE, AND AT EXTREMELY
CC LOW LEVELS (IF AT ALL) IN TESTIS AND BRAIN TISSUES.
CC -1- SIMILARITY: CONTAINS 1 LIM DOMAIN. THE LIM DOMAIN BINDS 2 ZINC
CC IONS.
CC -1- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U23769; AAA92046.1;
CC InterPro: IPR001781; LIM.
CC InterPro: IPR001478; PDZ.
CC Pfam: PF00412; LIM; 1.
CC Pfam: PF00595; PDZ; 1.
CC ProDom: PD000094; LIM; 1.
CC SMART: SM00132; LIM; 1.
CC SMART: SM00228; PDZ; 1.
CC PROSITE: PS00478; LIM_DOMAIN_1; 1.
CC PROSITE: PS50023; LIM_DOMAIN_2; 1.
CC PROSITE: PS50106; PDZ; 1.
CC LIM domain; Metal-binding; Zinc.
CC DOMAIN 1 84 PDZ.
CC FT DOMAIN 258 308 LIM.
CC SEQUENCE 327 AA; 35525 MW; 4C3B5F579FEDA4D1 CRC64;

Query Match 17.18; Score 118.5; DB 1; Length 327;
Best Local Similarity 39.24; Pred. NO. 0.00016;
Matches 31; Conservative 12; Mismatches 33; Indels 3; Gaps 2;

QY 32 LEGGAPWGFTLKGLEHGEPLIISKVEGKADTLSKLOAGDEVVHIN-EVTLSSSRKE 90
DB 8 LQPGWGFVGLVGKDFEQPLAISRYTPGSKA--AIANLCIGDLITAIKEDTSSMTHLE 65
QY 91 AVSLVKSGLRLLVRRN 109
DB 66 AONKIRKGVNDNMTLTVSRS 84

RESULT 6
RIL_RAT ID RIL_RAT STANDARD; PRT; 330 AA.
AC P36202;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE LIM PROTEIN RIL (REVERSION-INDUCED LIM PROTEIN) (RIT-18) (H-REV18).
GN RIL.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=Liver;
RX MEDLINE=95124724; PubMed=7824279;
RA Kiess M., Scharn B., Aguzzi A., Hajnal A., Klemenz R.,
RA Schwarte-Waldhoff I., Schaefer R.;
RT "Expression of ril, a novel LIM domain gene, is down-regulated in
RT Hras-transformed cells and restored in phenotypic revertants";
RL Oncogene 10:61-68(1995).
CC -1- TISSUE SPECIFICITY: DETECTED IN SEVERAL TISSUES, MOST PROMINENT IN
CC BRAIN AND HEART OF ADULTS.
CC -1- SIMILARITY: CONTAINS 1 LIM DOMAIN. THE LIM DOMAIN BINDS 2 ZINC
CC IONS.
CC -1- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: X76454; CAA53992.1;
CC PIR: S39246; S39246.
CC InterPro: IPR001781; LIM.
CC InterPro: IPR001478; PDZ.
CC Pfam: PF00412; LIM; 1.
CC Pfam: PF00595; PDZ; 1.
CC ProDom: PD000094; LIM; 1.
CC SMART: SM00132; LIM; 1.
CC SMART: SM00228; PDZ; 1.
CC PROSITE: PS00478; LIM_DOMAIN_1; 1.
CC PROSITE: PS50023; LIM_DOMAIN_2; 1.
CC PROSITE: PS50106; PDZ; 1.
CC LIM domain; Metal-binding; Zinc.
CC DOMAIN 1 84 PDZ.
CC FT DOMAIN 255 305 LIM.
CC SEQUENCE 330 AA; 35521 MW; 84BEB1E629ADA584 CRC64;

Query Match 16.8; Score 116.5; DB 1; Length 330;
Best Local Similarity 41.04; Pred. No. 0.00026;
Matches 32; Conservative 9; Mismatches 34; Indels 3; Gaps 2;

QY 32 LEGGAPWGFTLKGLEHGEPLIISKVEGKADTLSKLOAGDEVVHIN-EVTLSSSRKE 90
DB 7 LRGPSPWGFVGLVGKDFEQPLAISRYTPGSKA--ALAAALCPGDSIQAINGESTELMTHLE 64
QY 91 AVSLVKSGLRLLVRR 108

Db 65 AONRIKCHDLITLSVSR 82

```

RESULT 7
RIL_CHICK STANDARD: PRT: 330 AA.
ID RIL_CHICK
AC 09PW72;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE LIM PROTEIN RIL (REVERSION-INDUCED LIM PROTEIN).
GN RIL.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauromorpha; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
CC NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-White Leghorn;
RX MEDLINE=20379359; PubMed=10918612;
RA Fu S.-L., Maha A., Vogt P.R.;
RT Identification and characterization of genes upregulated in cells
transformed by v-Jun.
RL Oncogene 19:3537-3545(2000).
CC -1- FUNCTION: NEGATIVE REGULATOR OF CELL GROWTH.
CC -1- SIMILARITY: CONTAINS 1 LIM DOMAIN. THE LIM DOMAIN BINDS 2 ZINC
IONS.
CC -1- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.
CC CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: AF167295; AAD46655.1; -
CC DR InterPro: IPR001781; LIM.
CC DR InterPro: IPR001478; PDZ.
CC DR Pfam: PF00412; LIM; 1.
CC DR Pfam: PF00595; PDZ; 1.
CC DR ProDom: PD000094; LIM; 1.
CC DR SMART: SM00132; LIM; 1.
CC DR SMART: SM00228; PDZ; 1.
CC DR PROSITE: PS00478; LIM_DOMAIN_1; 1.
CC DR PROSITE: PS00023; LIM_DOMAIN_2; 1.
CC DR PROSITE: PS50106; PDZ; 1.
CC DR LIM domain; Metal-binding; zinc.
CC FT DOMAIN 8 84 PDZ.
CC FT DOMAIN 255 305 LIM.
CC SO SEQUENCE 330 AA; 35838 MW; 2C0E0B404301E404 CRC64;

Query Match 16.2%; Score 112.5; DB 1; Length 330;
Best Local Similarity 38.5%; Pred. No. 0.00065;
Matches 30; Conservative 12; Mismatches 33; Indels 3; Gaps 2;

```

```

DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PRESYNAPTIC PROTEIN SAP97 (SYNAPSE-ASSOCIATED PROTEIN 97) (DISCS,
DE LARGE HOMOLOG 1).
GN Dlg1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
CC NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95198112; PubMed=7891172;
RA Mueller B.M., Kistner U., Voh R.W., Cases-Langhoff C., Becker B.,
RA Gundelfinger E.D., Garner C.C.;
RT Molecular characterization and spatial distribution of SAP97, a
RT novel presynaptic protein homologous to SAP90 and the Drosophila
RT discs-large tumor suppressor protein.
RL J. Neurosci. 15:2354-2366(1995).
CC -1- FUNCTION: INTERACTS WITH THE CYTOPLASMIC TAIL OF NMDA RECEPTOR
CC SUBUNITS. MAY BE INVOLVED IN SYNAPTOGENESIS. MAY PLAY A ROLE IN
CC CELL ADHESION.
CC -1- SUBCELLULAR LOCATION: IN CULTURED T84 CELLS, IT IS RESTRICTED TO
CC THE CYTOPLASMIC SURFACE OF THE PLASMA MEMBRANES BETWEEN ADJACENT
CC CELLS, BUT NOT AT THE EDGES OF CELLS LACKING CELL-CELL CONTACT.
CC -1- TISSUE SPECIFICITY: PRESYNAPTIC NERVE TERMINI OF EXCITATORY
CC SYNAPSES. IN OTHER BRAIN REGIONS, SAP97 IS FOUND IN AND ALONG
CC BUNDLES OF UNMYELINATED AXONS. SAP97 IS NOT RESTRICTED TO THE CNS,
CC BUT IS ALSO PRESENT AT THE BASAL LATERAL MEMBRANE BETWEEN A
CC VARIETY OF EPITHELIAL CELLS.
CC -1- SIMILARITY: CONTAINS 3 PDZ/DHR DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -1- SIMILARITY: CONTAINS A GUANYLATE KINASE-LIKE DOMAIN.
CC -1- SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS.
CC CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: U14950; AAA79976.1; -
CC DR HSP: Q12958; IPDR.
CC DR InterPro: IPR000619; Guanylate_kin.
CC DR InterPro: IPR001478; PDZ.
CC DR InterPro: IPR001452; SH3.
CC DR Pfam: PF00625; Guanylate_kin; 1.
CC DR Pfam: PF00595; PDZ; 3.
CC DR Pfam: PF00018; SH3; 1.
CC DR SMART: SM00072; GUKC; 1.
CC DR SMART: SM00228; PDZ; 3.
CC DR SMART: SM00326; SH3; 1.
CC DR PROSITE: PS00856; GUANYLATE_KINASE_1; 1.
CC DR PROSITE: PS50052; GUANYLATE_KINASE_2; 1.
CC DR PROSITE: PS50106; PDZ; 3.
CC DR PROSITE: PS50002; SH3; 1.
CC DR SH3 domain; Repeat.
CC FT DOMAIN 224 310 PDZ 1.
CC FT DOMAIN 318 404 PDZ 2.
CC FT DOMAIN 465 545 PDZ 3.
CC FT DOMAIN 580 650 SH3.
CC FT DOMAIN 721 911 GUANYLATE KINASE.
CC FT DOMAIN 527 530 POLY-ALA.
CC SO SEQUENCE 911 AA; 100570 MW; 1CCEBD31DD0CAF8B CRC64;

Query Match 16.2%; Score 112.5; DB 1; Length 911;
Best Local Similarity 27.9%; Pred. No. 0.0021;
Matches 34; Conservative 27; Mismatches 42; Indels 19; Gaps 5;

```

QY 3 RTEDFKPSATLNSNTATKGRYIYL-EAFL-----EGAPWGTLLKGL 46
 Db 424 QSDNHVSPSSVLGQTPASPARSPISKAVLGDDSDITREPKVVLHRSSTGLGFINVGG- 482
 QY 47 EGEPLIISKVEEGKADTLSSKLAQAGDEVVHINEVTL-SSRKEAVSLVKSQKTLRLV 105
 Db 483 EDGEGIFISFILAGPAD-LSGELRKGDRIIISVNSVDLRAASHEQAAAAALNAGQAVTI 541
 QY 106 VR 107
 Db 542 AQ 543

RESULT 9
 PTN4_HUMAN STANDARD; PRT; 926 AA.
 ID PTN4_HUMAN
 AC P29074;
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 01-DEC-1992 (Rel. 24, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE PROTEIN TYROSINE PHOSPHATASE, NON-RECEPTOR TYPE 4 (EC 3.1.3.48)
 DE (PROTEIN-TYROSINE PHOSPHATASE MEG1) (PTPASE-MEG1) (MEG).
 GN PTPN4.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 [1]
 SEQUENCE FROM N.A.
 RX MEDLINE-91288564; Pubmed-1648233;
 RA Gu M., York J.D., Warshawsky I., Majerus P.W.;
 RT Identification, cloning, and expression of a cytosolic megakaryocyte
 RT protein-tyrosine-phosphatase with sequence homology to cytoskeletal
 RT protein 4.1.;
 RL Proc. Natl. Acad. Sci. U.S.A. 88:5867-5871(1991).
 CC -1- FUNCTION: MAY ACT AT JUNCTIONS BETWEEN THE MEMBRANE AND THE
 CC CYTOSKELETON.
 CC -1- CATALYTIC ACTIVITY: PROTEIN TYROSINE PHOSPHATE + H(2O) -
 CC PROTEIN TYROSINE + ORTHOPHOSPHATE.
 CC -1- SIMILARITY: CONTAINS 1 BAND 4.1-LIKE DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.
 CC -1- SIMILARITY: BELONGS TO THE NON-RECEPTOR CLASS OF THE PROTEIN-
 CC TYROSINE PHOSPHATASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M68941; AAA36530.1; -
 CC PIR; A41105; A41105.
 CC HSSP; P18052; 1YFO.
 CC MIN; 176878; -
 CC InterPro; IPR000299; Band_4.1.
 CC InterPro; IPR001478; PDZ.
 CC InterPro; IPR000387; Tyr_phosphatase.
 CC InterPro; IPR000242; Tyr_prot_phphatase.
 CC Pfam; PF00373; Band_4.1; 1.
 CC Pfam; PF00595; PDZ; 1.
 CC Pfam; PF0102; Y_phosphatase; 1.
 CC PRINTS; PR00700; PRTYPHPTASE.
 CC PRINTS; PR00935; BAND4.1.
 CC SMART; SM00295; B41; 1.
 CC SMART; SM00228; PDZ; 1.
 CC SMART; SM00194; PTPC; 1.
 CC PROSITE; PS00660; BAND_4.1; 1.
 CC PROSITE; PS00661; BAND_4.1.2; 1.
 CC PROSITE; PS50057; BAND_4.1.3; 1.
 CC PROSITE; PS50106; PDZ; 1.
 CC PROSITE; PS00383; TYR_PHOSPHATASE_1; 1.

DR PROSITE; PS50056; TYR_PHOSPHATASE_2; 1.
 DR PROSITE; PS50055; TYR_PHOSPHATASE_PTP; 1.
 KW Structural protein; Cytoskeleton; Hydrolase.
 FT DOMAIN 84 241 BAND 4.1-LIKE.
 FT DOMAIN 517 589 PDZ.
 FT DOMAIN 677 926 PROTEIN-TYROSINE PHOSPHATASE.
 FT ACT_SITE 852 852 BY SIMILARITY.
 SQ SEQUENCE 926 AA; 105911 MW; 4DAC6A87A675CFB0 CRC64;
 Query Match 16.2%; Score 112.5; DB 1; Length 926;
 Best Local Similarity 30.0%; Pred. No. 0.0022;
 Matches 33; Conservative 22; Mismatches 36; Indels 19; Gaps 4;
 QY 38 WGFTLKGLEHGEPLIISKVEEGKADTLSSKLAQAGDEVVHINEVTLSS-SRKEAVSLVK 96
 Db 528 FGFNVKGGYDQKMPVIVSRVAPGTPADLCVPRLNQEGDVVLINGRDIAETHDQVVLFIK 587
 QY 97 GSYK-----TLRLVVRNGV-----LLCQPEWKV-----AIHRDH 128
 Db 588 ASCERHSGELMLLRPNNAVYDVVVEKLENPDFQYIPEKAPLDSVHQDDH 637
 RESULT 10
 DLG3_MOUSE STANDARD; PRT; 849 AA.
 ID DLG3_MOUSE
 AC P70175;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE PRESYNAPTIC PROTEIN SAPI02 (SYNAPSE-ASSOCIATED PROTEIN 102) (DISCS,
 DE LARGE HOMOLOG 3).
 GN DLG3 OR DLGH3.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 OX NCBI_TaxID=10090;
 [1]
 SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6; TISSUE=Brain;
 RA Kohmura N., Makino S., Yagi T.;
 RL Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: INTERACTS WITH THE CYTOPLASMIC TAIL OF THE NMDA RECEPTOR
 CC SUBUNIT NR2B.
 CC -1- SIMILARITY: CONTAINS 3 PDZ/DHR DOMAINS.
 CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
 CC -1- SIMILARITY: CONTAINS A GUANYLATE KINASE-LIKE DOMAIN.
 CC -1- SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; D87117; BAAL3249.1; -
 CC HSSP; Q12959; IPDR.
 CC MGD; MGI:1888986; Digh3.
 CC InterPro; IPR000619; Guanylate_kin.
 CC InterPro; IPR001478; PDZ.
 CC InterPro; IPR001452; SH3.
 CC Pfam; PF00625; Guanylate_kin; 1.
 CC Pfam; PF00595; PDZ; 3.
 CC Pfam; PF00018; SH3; 1.
 CC SMART; SM00072; GUKC; 1.
 CC SMART; SM00228; PDZ; 3.
 CC SMART; SM00326; SH3; 1.
 CC PROSITE; PS00856; GUANYLATE_KINASE_1; 1.
 CC PROSITE; PS50052; GUANYLATE_KINASE_2; 1.
 CC PROSITE; PS50106; PDZ; 3.
 CC PROSITE; PS50002; SH3; 1.

```

DR PFam; PF00006; SH3; 1.
DR SMART; SM00072; GUKC; 1.
DR SMART; SM00328; PDZ; 3.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS00856; GUANYLATE_KINASE_1; 1.
DR PROSITE; PS50052; GUANYLATE_KINASE_2; 1.
DR PROSITE; PS50106; PDZ; 3.
DR PROSITE; PS50002; SH3; 1.
DR SH3 domain; Repeat; Alternative splicing.
KM DOMAIN 149 235 PDZ 1.
FT DOMAIN 244 330 PDZ 2.
FT DOMAIN 404 484 PDZ 3.
FT DOMAIN 519 589 SH3.
FT DOMAIN 659 849 GUANYLATE_KINASE.
FT VASPLIC 627 640 MISSING (IN SHORT ISOFORM).
SQ SEQUENCE 849 AA; 93539 MW; 34DA9C46C7B96DB CRC64;

Query Match 16.0%; Score 111.5; DB 1; Length 849;
Best Local Similarity 33.0%; Pred. No. 0.0025;
Matches 35; Conservative 23; Mismatches 37; Indels 11; Gaps 6;

Qy 10 KPSATLNSNTATKGRITYLEAFLE-GGAPNPFITLKGLEH-----GEP-LIISKVEGSGKA 63
Db 133 EPSLNV--GSDGEMFKEEIYLENGNSGLGFSIAGGIDNHPDDPGIFITKLIPEGAA 189
Qy 64 DPLSSKLAGDEVYHINEYTLSS-SRKAVSLVKSRYETRLVYVR 108
Db 190 -AMDRLGVNDCVLAVNEVDYSEVYHSHRAVELKAGFVVALVYVR 234

RESULT 12
DGL1_HUMAN
AC Q12959; Q12958; STANDARD; PRT; 904 AA.
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PRESNAPTIC PROTEIN SAP97 (SYNAPSE-ASSOCIATED PROTEIN 97) (DISCS,
DE LARGE HOMOLOG 1).
GN DGL1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95024052; PubMed=7937897;
RA Lue R.A., Marfatia S.M., Branton D., Chishti A.H.;
RT "Cloning and characterization of hdlg, the human homologue of the
RT Drosophila discs large tumor suppressor binds to protein 4.1.";
RT Proc. Natl. Acad. Sci. U.S.A. 91:9818-9822(1994).
RN [3]
RX X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS) OF 460-555.
RA MEDLINE=96388231; PubMed=8757139;
RA Cabral J.H., Petosa C., Sutcliffe M.J., Raza S., Byron O., Poy F.,
RA Marfatia S.M., Chishti A.H., Liddington R.C.;
RT "Crystal structure of a PDZ domain.";
RT Nature 382:649-652(1996).
RN [1]
CC -1 FUNCTION: INTERACTS WITH THE CYTOPLASMIC TAIL OF NMDA RECEPTOR
CC SUBUNIT: ASSOCIATES WITH PROTEIN 4.1.
CC -1 ALTERNATIVE PRODUCTS: 2 ISOFORMS; 1 (SHOWN HERE) AND 2; ARE
CC PRODUCED BY ALTERNATIVE SPLICING.
CC -1 SIMILARITY: CONTAINS 3 PDZ/DHR DOMAINS.
CC -1 SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -1 SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```

CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).

CC EMBL; U13897; AAA50599.1; -
 CC EMBL; U13896; AAA50598.1; -
 DR PDB; 1PDR; 23-JUL-97.
 DR MIM; 601014; -
 DR InterPro; IPR000619; Guanylate_kin.
 DR InterPro; IPR001478; PDZ.
 DR InterPro; IPR001452; SH3.
 DR Pfam; PF00625; Guanylate_kin; 1.
 DR Pfam; PF00595; PDZ; 3.
 DR Pfam; PF00018; SH3; 1.
 DR SMART; SM00072; GUKC; 1.
 DR SMART; SM00228; PDZ; 3.
 DR SMART; SM00326; SH3; 1.
 DR PROSITE; PS00856; GUANYLATE_KINASE_1; 1.
 DR PROSITE; PS00052; GUANYLATE_KINASE_2; 1.
 DR PROSITE; PS0106; PDZ; 3.
 DR PROSITE; PS00002; SH3; 1.
 KW SH3 domain; Repeat; Alternative splicing; 3D-structure.
 FT DOMAIN 224 310 PDZ 1.
 FT DOMAIN 319 405 PDZ 2.
 FT DOMAIN 466 546 PDZ 3.
 FT DOMAIN 581 651 SH3.
 FT DOMAIN 714 904 GUANYLATE KINASE.
 FT VARSPPLIC 669 680 EIPDDMGSKGLK -> QSFNDRKKNLFSRPFYKNDQS
 FT EIPDDMGSKGLK (IN ISOFORM 2).
 FT EIPDDMGSKGLK (IN ISOFORM 2).
 FT SEQUENCE 904 AA; 100354 MW; B78798D6B0920D4 CRC64;

Query Match 16.0%; Score 111; DB 1; Length 904;
 Best Local Similarity 30.3%; Pred. No. 0.003;
 Matches 36; Conservative 23; Mismatches 40; Indels 20; Gaps 6;
 QY 7 DFF-KPSATLNSNTAKGRYIVL-BAFL-----EGGAPWGFTLKGLGHEG 49
 Db 428 DNHVSFSLGQTSPSPARYSPVSRVAVLGDDEITREPKVYLHRSSTGLGFNIVGG-EDG 486
 QY 50 EPLIISKVEGKADTLSSKLQAGDEVHNEVTL-SSSRKEAVSLVSGSYKTLRLVVR 107
 Db 487 EGIFISFILLAGGPAD-LSGELRKGRDRIISVNSVDLRAASHEQAAALNAGQAVTIAQ 544

RESULT 13
 PDLL_HUMAN
 ID PDLL_HUMAN STANDARD; PRT; 329 AA.
 AC 000151;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE PDZ AND LIM DOMAIN PROTEIN 1 (LIM DOMAIN PROTEIN CLP-36) (C-TERMINAL
 DE LIM DOMAIN PROTEIN 1) (ELFIN).
 GN PDLLIM OR CLIMI OR CLP36.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Heart;
 RX MEDLINE=99145108; PubMed=10022510;
 RA Kotaka M., Ngai S.M., Garcia-Barcelo M., Tsui S.K.W., Fung K.P.,
 RA Lee C.Y., Waye M.M.Y.;
 RT "Characterization of the human 36-kDa carboxyl terminal LIM domain
 RT protein (hCLIM1)."
 RL J. Cell. Biochem. 72:279-285(1999).
 RN [2]
 RP INTERACTION WITH ALPHA-ACTININ 2.
 RX MEDLINE=20320865; PubMed=10861853;
 RA Kotaka M., Kustin S., Ngai S., Chan K., Lau Y., Lee S.M., Li H.Y.,
 RA Ng E.K., Schaper J., Tsui S.K.W., Fung K.P., Lee C.Y., Waye M.M.Y.;
 RT "Interaction of hCLIM1, an enigma family protein, with alpha-actinin

RT 2.;
 RL J. Cell. Biochem. 78:558-565(2000).
 RN [3]
 RP INTERACTION WITH ALPHA-ACTININS 1 AND 4.
 RX MEDLINE=20219155; PubMed=10753915;
 RA Vallienius T., Luukko K., Makela T.P.;
 RT "CLP-36 PDZ-LIM protein associates with nonmuscle alpha-actinin-1 and
 RT alpha-actinin-4."
 RL J. Biol. Chem. 275:11100-11105(2000).
 CC -1- FUNCTION: CYTOSKELETAL PROTEIN THAT MAY ACT AS AN ADAPTER THAT
 CC BRINGS OTHER PROTEINS TO THE CYTOSKELETON.
 CC -1- SUBUNIT: INTERACTS WITH ALPHA-ACTININS, 1 2 AND 4.
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
 CC -1- TISSUE SPECIFICITY: STRONGLY EXPRESSED IN THE HEART AND SKELETAL
 CC MUSCLE, MODERATELY EXPRESSED IN THE LIVER, SMALL INTESTINE,
 CC COLON, PLACENTA, AND LUNG. A LOWER LEVEL EXPRESSION IS SEEN IN
 CC LIVER, THYMUS, KIDNEY, PROSTATE AND PANCREAS AND IS NOT FOUND IN
 CC THE BRAIN, TESTIS, OVARY, AND PERIPHERAL BLOOD LEUKOCYTES.
 CC -1- SIMILARITY: CONTAINS 1 LIM DOMAIN. THE LIM DOMAIN BINDS 2 ZINC
 CC IONS.
 CC -1- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).

EMBL; U90878; AAC05580.1; -
 MIM; 605900; -
 DR InterPro; IPR001781; LIM.
 DR InterPro; IPR001478; PDZ.
 DR Pfam; PF00412; LIM; 1.
 DR Pfam; PF00595; PDZ; 1.
 DR ProDom; PD000094; LIM; 1.
 DR SMART; SM00132; LIM; 1.
 DR SMART; SM00228; PDZ; 1.
 DR PROSITE; PS00478; LIM_DOMAIN_1; FALSE_NEG.
 DR PROSITE; PS00023; LIM_DOMAIN_2; 1.
 DR PROSITE; PS0106; PDZ; 1.
 KW Cytoskeleton; LIM domain; Metal-binding; Zinc.
 FT DOMAIN 3 84 PDZ.
 FT DOMAIN 260 310 LIM.
 FT SEQUENCE 329 AA; 36171 MW; 6520F1A932CFA312 CRC64;

Query Match 15.8%; Score 109.5; DB 1; Length 329;
 Best Local Similarity 38.0%; Pred. No. 0.0013;
 Matches 30; Conservative 12; Mismatches 34; Indels 3; Gaps 2;
 QY 32 LEGGAPWGFTLKGLGHEGEPLIISKVEGGRADTLSSKLQAGDEVHIN-EVTLSSSRKE 90
 Db 8 LQGPCPGWFRVLGRKDFEQPLAISRVTPGSKA--ALANLCIGDVTADIGTNTMTHLE 65
 QY 91 AVSLVKGYSYKTLRLVVRN 109
 Db 66 AQNRKIGCTDNLTLTVARS 84
 RESULT 14
 DLG4_MOUSE
 ID DLG4_MOUSE STANDARD; PRT; 724 AA.
 AC 062108;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE PRESYNAPTIC DENSITY PROTEIN 95 (PSD-95) (PRESYNAPTIC PROTEIN SAP90)
 DE (SYNAPSE-ASSOCIATED PROTEIN 90) (DISCS, LARGE HOMOLOG 4).
 GN DLG4 OR DLGH4 OR PSD95.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;


```

FT DOMAIN 160 246 PDZ 2.
FT DOMAIN 313 393 PDZ 3.
FT DOMAIN 428 498 SH3
FT DOMAIN 534 724 GUANYLATE KINASE.
FT CONFLICT 61 61 M -> L (IN REF. 2).
FT CONFLICT 78 78 S -> T (IN REF. 2).
FT CONFLICT 177 182 GVGNOH -> ALGTSI (IN REF. 2).
FT CONFLICT 200 200 A -> G (IN REF. 2).
FT CONFLICT 254 254 S -> T (IN REF. 2).
FT CONFLICT 539 555 ILGPTKDRANDLLSEF -> ISLDPPKTVPTMIFSPSS
(IN REF. 2).
FT CONFLICT 523 625 GKX -> RDQ (IN REF. 3).
SQ SEQUENCE 724 AA; 80465 MW; 7922D4E8E0F9AD85 CRC64;

Query Match 15.08; Score 104.5; DB 1; Length 724;
Best Local Similarity 34.7%; Pred. No. 0.01;
Matches 26; Conservative 20; Mismatches 26; Indels 3; Gaps 3;

QY 34 GGAPWGFTLKGLGEPLIISKVEEGKADTSLSKLQAGDEVVHINEVTI-SSSRKEAV 92
| : | | | | : | | | | | : | | : | | : | | : | | : | | : |
Db 319 GSTGLGFNIVGG-EDGEGIFISFILAGGPAD-LSGELRKGDQILSVNGVDLRNASHEQAA 376
| : | | | | : | | | | | : | | : | | : | | : | | : | | : |

QY 93 SLVKGSYKTLRLVLR 107
| : | | : | :
Db 377 IALKNAGQTVTIIAQ 391

```

Search completed: January 31, 2002, 21:06:29
Job time: 52 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 21:05:35 ; Search time 72.88 Seconds
(without alignments)
266,935 Million cell updates/sec

Title: US-09-641-831-6
Perfect score: 695
Sequence: 1 MKRTTDDHFKPSATLNSNTA.....CCPEKVVAIHRHDHALQP 133

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 200000000
Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: SPREMBL.17:*
2: sp_archaea:*
3: sp_bacteria:*
4: sp_fungi:*
5: sp_human:*
6: sp_invertebrate:*
7: sp_mammal:*
8: sp_mhc:*
9: sp_organelle:*
10: sp_phage:*
11: sp_plant:*
12: sp_rodent:*
13: sp_virus:*
14: sp_vertebrate:*
15: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	415	59.7	1986	11	Q9GXNO mus musculus
2	273	39.3	1502	4	Q9ULI8 homo sapien
3	145	20.9	1454	4	Q9H4G2
4	134.5	19.4	298	13	Q9PU46
5	132.5	19.1	315	13	Q9PU47
6	130	18.7	493	5	Q9W443
7	128	18.4	39	4	Q9H4T0
8	128	18.4	403	4	Q9UPC3
9	128	18.4	548	11	Q9ES65
10	128	18.4	552	4	Q9UM04
11	128	18.4	652	4	Q9Y6N9
12	128	18.4	910	5	Q9ES64
13	127.5	18.3	400	11	Q9V7F2
14	127.5	18.3	1171	11	Q9A893
15	126	18.1	521	4	Q9UM17
16	124.5	17.9	207	4	Q9H4P6
17	123.5	17.8	207	11	Q9Z252
18	123.5	17.8	207	11	Q9H951
19	123.5	17.8	234	11	Q9D2Q1

20	123.5	17.8	1256	4	Q75085	Q75085 homo sapien
21	123	17.7	171	5	Q17458	Q17458 schistosoma
22	122.5	17.6	596	4	Q60705	Q60705 homo sapien
23	121	17.4	202	4	Q9NTP3	Q9NTP3 homo sapien
24	121	17.4	517	4	Q9H5P4	Q9H5P4 homo sapien
25	120	17.3	926	11	Q9W022	Q9W022 mus musculus
26	118.5	17.1	138	11	Q9D130	Q9D130 mus musculus
27	118.5	17.1	288	11	Q9R021	Q9R021 mus musculus
28	118.5	17.1	288	11	Q9WVH1	Q9WVH1 mus musculus
29	117.5	16.9	199	4	Q9BVP4	Q9BVP4 homo sapien
30	117.5	16.9	214	11	Q9CRA2	Q9CRA2 mus musculus
31	117.5	16.9	239	11	Q9QYNI	Q9QYNI mus musculus
32	117.5	16.9	315	13	Q98RT5	Q98RT5 salmo salar
33	117.5	16.9	327	11	Q98K93	Q98K93 mus musculus
34	117.5	16.9	329	4	Q9BVP2	Q9BVP2 homo sapien
35	117.5	16.9	337	11	Q9QYNO	Q9QYNO mus musculus
36	117.5	16.9	364	4	Q43580	Q43580 homo sapien
37	117.5	16.9	591	11	Q62920	Q62920 rattus norv
38	117.5	16.9	591	11	Q9QYNI	Q9QYNI mus musculus
39	116.5	16.8	283	4	Q9Y4Z5	Q9Y4Z5 homo sapien
40	116.5	16.8	470	4	Q9Y4Z3	Q9Y4Z3 homo sapien
41	116.5	16.8	617	4	Q9Y4Z4	Q9Y4Z4 homo sapien
42	116.5	16.8	661	11	Q9JRS3	Q9JRS3 mus musculus
43	116.5	16.8	723	11	Q9JRS4	Q9JRS4 mus musculus
44	116.5	16.8	723	11	Q9WVH2	Q9WVH2 mus musculus
45	116.5	16.8	734	4	Q75112	Q75112 homo sapien

ALIGNMENTS

RESULT 1
ID Q9GXNO PRELIMINARY: PRT: 1986 AA.
AC Q9GXNO:
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE PDZ DOMAIN ACTIN BINDING PROTEIN SHROOM.
GN SHRM.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20055594; PubMed=10589677;
RA Hildebrand J.D., Soriano P.;
RT "Shroom, a PDZ domain-containing actin-binding protein, is required
RT for neural tube morphogenesis in mice.";
RL Cell 99:485-497(1999).
DR EMBL: AF199421; AAF13269.1; -
DR HSPF: P29476; IOAV.
DR WGD: MG1:151655; Shrm.
DR InterPro: IPR001478; PDZ.
DR Pfam: PF00595; PDZ; 1.
DR SMART: SM00228; PDZ; 1.
DR PROSITE: PS50106; PDZ; 1.
SQ SEQUENCE 1986 AA: 215261 MW: 18056FCBB772EABF CRC64;

Query Match 59.7%; Score 415; DB 11; Length 1986;
Best local similarity 75.0%; Pred. No. 8.3e-31;
Matches 81; Conservative 12; Mismatches 15; Indels 0; Gaps 0;
QY 2 MKRTTDDHFKPSATLNSNTATKGRVITLFAFLBGGAGWGTLLKGLGHRGPRLLISKYEEGG 61
1 MKRTTDDHFKPSATLNSNTATKGRVITLFAFLBGGAGWGTLLKGLGHRGPRLLISKYEEGG 60
QY 62 KATLTSKIQAGEVYHNEVTLSSRKAVSLVKSYSYTLRLVVRN 109
61 KADSVSSGLOAGEVYHNEVTLSSRKAVSLVKSYSYTLRLVVRN 108
Db

Fri Feb 1 08:56:37 2002

RESULT 2
Q9ULL8 PRELIMINARY; PRT; 1502 AA.
AC Q9ULL8;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE KIAA1202 PROTEIN (FRAGMENT).
GN KIAA1202.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP TISSUE=BRAIN;
RC MEDLINE=20039619; PubMed=10574462;
RA Nagase T., Ishikawa K., Kikuno R., Hirose M., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XV.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 6:337-345(1999).
DR EMBL; AB033028; BAA86516.1; -.
DR HSSP; P31016; IBE9.
DR InterPro; IPR001478; PDZ.
DR Pfam; PF00595; PDZ; 1.
DR SMART; SM00228; PDZ; 1.
DR PROSITE; PS50106; PDZ; 1.
DR NON_TER 1
FT SEQUENCE 1502 AA; 165950 MW; DC050C9C69B78AD9 CRC64;

Query Match 39.38; Score 273; DB 4; Length 1502;
Best Local Similarity 52.98; Pred. No. 2.2e-17;
Matches 54; Conservative 19; Mismatches 27; Indels 2; Gaps 2;

QY 23 GRVYILEAFLEGAPWGFTLKGLEHGEPLIISKVEGGKADTLSSKIQAGDEVVHINEV 82
DB 10 GSFQVQVQGGAPWGFTLKGLEHGEPLIISKVEGGKADTLSSKIQAGDEVVHINEV 68

QY 83 TLSSSRKEAVSLVKGSKYTLRLVVRNGVLLCCP-EWKVVAI 123
DB 69 PLYGSRQEAALIKGSRFLKLVRRNAPVSRPHSHVAKL 110

RESULT 3
Q9H4G2 PRELIMINARY; PRT; 1454 AA.
AC Q9H4G2;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE BAI19E20.1 (KIAA1202 PROTEIN) (FRAGMENT).
GN BAI19E20.1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP TISSUE=BRAIN;
RC MEDLINE=20039619; PubMed=10574462;
RA Nagase T., Ishikawa K., Kikuno R., Hirose M., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XV.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 6:337-345(1999).
DR EMBL; AB033028; BAA86516.1; -.
DR HSSP; P31016; IBE9.
DR InterPro; IPR001478; PDZ.
DR Pfam; PF00595; PDZ; 1.
DR SMART; SM00228; PDZ; 1.
DR PROSITE; PS50106; PDZ; 1.
DR NON_TER 1
FT SEQUENCE 1454 AA; 160615 MW; 7931571D7DA80D88 CRC64;

Query Match 20.98; Score 145; DB 4; Length 1454;
Best Local Similarity 45.68; Pred. No. 3.8e-05;

Matches 31; Conservative 15; Mismatches 20; Indels 2; Gaps 2;
QY 57 VEEGKADTLSSKIQAGDEVVHINEVTLSSSRKEAVSLVKGSKYTLRLVVRNGVLLCCP 116
DB 1 IEDGKGA-ALSKMRTGDELVNINGTPLYGSRQEAALIKGSRFLKLVRRNAPVSRP 59

QY 117 -EWKVVAI 123
DB 60 HSHVAKL 67

RESULT 4
Q9PU46 PRELIMINARY; PRT; 298 AA.
AC Q9PU46;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE ALPHA-ACTININ ASSOCIATED LIM PROTEIN, SKELETAL MUSCLE ISOFORM (SKALP) (FRAGMENT).
GN ALP.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP TISSUE=MUSCLE;
RC MEDLINE=99436131; PubMed=10506181;
RA Pomies P., Macalma T., Beckerle M.C.;
RT "Purification and characterization of an alpha-actinin-binding PDZ-LIM
RT protein that is up-regulated during muscle differentiation.";
RL J. Biol. Chem. 274:29242-29250(1999).
DR EMBL; AJ249219; CAB53971.1; -.
DR InterPro; IPR001478; PDZ.
DR Pfam; PF00595; PDZ; 1.
DR PROSITE; PS50106; PDZ; 1.
DR SMART; SM00228; PDZ; 1.
DR NON_TER 298
FT SEQUENCE 298 AA; 32094 MW; 27D6F24FC6141E61 CRC64;

Query Match 19.48; Score 134.5; DB 13; Length 298;
Best Local Similarity 36.88; Pred. No. 5.1e-05;
Matches 32; Conservative 17; Mismatches 35; Indels 3; Gaps 2;

QY 32 LEGAPWGFTLKGLEHGEPLIISKVEGGKADTLSSKIQAGDEVVHINEV-TLSSSRKE 90
DB 7 LPGPAPWGFRSLGGIDFNQPLIITRITPGSKAST--ANLCPGDIIVAINGLSTENWTHND 64

QY 91 AVSLVKGSKYTLRLVVRNGVLLCCPE 117
DB 65 AQRKAAAHQSLSLRIKRAETKLMSQ 91

RESULT 5
Q9PU47 PRELIMINARY; PRT; 315 AA.
AC Q9PU47;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE ALPHA-ACTININ ASSOCIATED LIM PROTEIN, SMOOTH MUSCLE ISOFORM (SMALP).
GN ALP.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP TISSUE=MUSCLE;
RC MEDLINE=99436131; PubMed=10506181;
RA Pomies P., Macalma T., Beckerle M.C.;

Purification and characterization of an alpha-actinin-binding PDZ-LIM protein that is up-regulated during muscle differentiation.
 RT J Biol. Chem. 274:29242-29250(1999).
 CC -1- SIMILARITY: CONTAINS LIM DOMAIN(S). THE LIM DOMAIN BINDS 2 ZINC IONS.
 CC EMBL: AJ249218; CAB53970.1; -
 DR InterPro: IPR001781; LIM.
 DR InterPro: IPR001478; PDZ.
 DR Pfam: PF00412; LIM; 1.
 DR Pfam: PF00595; PDZ; 1.
 DR ProDom: PD000094; LIM; 1.
 DR SMART: SM00132; LIM; 1.
 DR SMART: SM00228; PDZ; 1.
 DR PROSITE: PS50023; LIM_DOMAIN_2; 1.
 DR PROSITE: PS50106; PDZ; 1.
 KM LIM motif: Metal-binding; Zinc.
 SO SEQUENCE 315 AA; 34365 MW; 86414A0A89BCA35 CRC64;

Query Match 19.1%; Score 132.5; DB 13; Length 315;
 Best Local Similarity 36.8%; Pred. No. 8.6e-05;
 Matches 32; Conservative 16; Mismatches 36; Indels 3; Gaps 2;

QY 32 LEGAPMGFTLKGLEHGEPLISVEEGKADTLSSKQAGDEVYHINEY-TLSSSKRE 90
 Db 7 LPGAPEWGFRLSGIDENQPLITRTPGSKAST--ANLCREDITVAINGLSTENWTHND 64

QY 91 AVSLVKGSTKTLRLVVRNGVLLCPE 117
 Db 65 AOEKIKAAHQLSLERAEKTKLMSPO 91

RESULT 6
 Q9M443 PRELIMINARY; PRT; 493 AA.

AC Q9M443.01-MAR-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE CG5921.1 PROTEIN.
 GN CG5921.
 OS Drosophila melanogaster (Fruit fly).
 CC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 CC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 CC Ephydroidea; Drosophilidae; Drosophila.
 CC NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amaniatis P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.H.C., Heit G., Nelson C.R., Miklos G.L.G.,
 RA Wan K.H., Doyle C., Baxter E.G., Heit G., Nelson C.R., Miklos G.L.G.,
 RA Abail J.F., Abdayani A., An H.-J., Andrews-Frankoch C., Baldwin D.,
 RA Baller R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
 RA Burris K.C., Busan D.A., Butler H., Caddieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos J.M., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glöck A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman J.T., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegun C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasno P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,

RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeter F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Stryker R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao O.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster";
 RL Science 287:2185-2195(2000).
 DR EMBL: AE003436; AAF46116.1; -
 DR HSSP: Q12959; 1PDR.
 DR FlyBase: FBgn0029835; CG5921.
 DR InterPro: IPR001478; PDZ.
 DR Pfam: PF00595; PDZ; 2.
 DR SMART: SM00228; PDZ; 2.
 DR PROSITE: PS50106; PDZ; 2.
 SO SEQUENCE 493 AA; 50574 MW; C46711C9F19797A4 CRC64;

Query Match 18.7%; Score 130; DB 5; Length 493;
 Best Local Similarity 37.8%; Pred. No. 0.00026;
 Matches 37; Conservative 16; Mismatches 35; Indels 10; Gaps 5;

QY 33 EGAPMGFTLKGLEHGEPLISKEEGKADTLSSKQAGDEVYHINEYTLSSS-KREA 91
 Db 80 ERGSTYGFVTRGREGHGFVSHVEHGGAHL--KGIRIGDQILIRINGRLDAVHKEF 137

QY 92 VSLVKGSTKTLRLVVRNGVLLC--CPE---WKYVAI 123
 Db 138 IOLVAGQDR-VTLKVGVGMPLVRDLPERLSWSYVKL 174

RESULT 7
 Q9H4T0 PRELIMINARY; PRT; 39 AA.

AC Q9H4T0.01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
 DE BA554P16.1 (KIAA1202) (FRAGMENT).
 GN BA554P16.1.
 OS Homo sapiens (Human).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 CC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC Bird C.;
 RL Submitted (SEP-2000) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AL359272; CAC12948.1; -
 FT NON_TER 39
 SO SEQUENCE 39 AA; 4260 MW; 2773DAE613D25D3F CRC64;

Query Match 18.4%; Score 128; DB 4; Length 39;
 Best Local Similarity 67.6%; Pred. No. 1.6e-05;
 Matches 23; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

QY 23 GRTYVLEAFLEGAPMGFTLKGLEHGEPLISK 56
 Db 6 GSFQYVPVOJOGAPMGFTLKGLEHCEPLTVSK 39

RESULT 8
 Q9UPC3 PRELIMINARY; PRT; 403 AA.

ID Q9UPC3.09UPC3.
 AC Q9UPC3.

Fri Feb 1 08:56:37 2002

```
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE ANTIGEN NY-CO-37.
GN NY-CO-38.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-COLON CANCER METASTASIS TO LUNG;
RX MEDLINE=9627252; PubMed=9610721;
RA Scanlan M.J., Chen Y.T., Williamson B., Gure A.O., Stockert E.,
RA Gordon J.D., Tureci O., Sahin U., Pfreundschuh M., Old L.J.;
RT "Characterization of human colon cancer antigens recognized by
RT autologous antibodies.";
RL Int. J. Cancer 76:652-658(1998).
DR EMBL; AF039699; AAC18048.1; .
DR HSSP; P29476; 1QAV.
DR InterPro; IPR001478; PDZ.
DR Pfam; PF00595; PDZ; 2.
DR SMART; SM00228; PDZ; 2.
DR PROSITE; PS0106; PDZ; 2.
SQ SEQUENCE 403 AA; 45838 MW; AB24FB6E1D3144FD CRC64;

Query Match 18.4%; Score 128; DB 4; Length 403;
Best Local Similarity 40.0%; Pred. No. 0.00032;
Matches 30; Conservative 22; Mismatches 19; Indels 4; Gaps 3;

QY 39 GFTLKGLEGEPLISKVEGGKADTLSSKLQAGDEVVHINEVTLS-SRKEAVSLVK 97
DB 99 GLSVRGLEFGCGLFISHIKGGQADSVG--LQVGDEIVRINGYSISCTHEEVINLIR- 155

QY 98 SYKTLRLVVRNGVL 112
DB 156 TKKTVSIKVRHIGLI 170

RESULT 9
QY9S65 PRELIMINARY; PRT; 548 AA.
AC Q9S65;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE HARMONIN ISOFORM A1.
GN 2010016F01RIK OR USH1C.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20428180; PubMed=10973247;
RA Verpy E., Leibovici M., Zwaenepoel I., Liu X.Z., Gal A., Salem N.,
RA Mansour A., Blanchard S., Kobayashi I., Keats B.J., Slim R., Petit C.;
RT "A defect in harmonin, a PDZ domain-containing protein expressed in
RT the inner ear sensory hair cells, underlies Usher syndrome type 1C.";
RL Nat. Genet. 26:51-55(2000).
DR EMBL; AF228924; AAC12457.1; .
DR MGD; MGI:1919338; 2010016F01RIK.
DR InterPro; IPR001478; PDZ.
DR Pfam; PF00595; PDZ; 3.
DR SMART; SM00228; PDZ; 3.
DR PROSITE; PS0106; PDZ; 3.
SQ SEQUENCE 548 AA; 61712 MW; 44A245AA32091070 CRC64;

Query Match 18.4%; Score 128; DB 11; Length 548;
Best Local Similarity 40.0%; Pred. No. 0.00047;
Matches 30; Conservative 22; Mismatches 19; Indels 4; Gaps 3;

DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE ANTIGEN NY-CO-37.
GN NY-CO-38.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-COLON CANCER METASTASIS TO LUNG;
RX MEDLINE=9627252; PubMed=9610721;
RA Scanlan M.J., Chen Y.T., Williamson B., Gure A.O., Stockert E.,
RA Gordon J.D., Tureci O., Sahin U., Pfreundschuh M., Old L.J.;
RT "Characterization of human colon cancer antigens recognized by
RT autologous antibodies.";
RL Int. J. Cancer 76:652-658(1998).
DR EMBL; AF039699; AAC18048.1; .
DR HSSP; P29476; 1QAV.
DR InterPro; IPR001478; PDZ.
DR Pfam; PF00595; PDZ; 2.
DR SMART; SM00228; PDZ; 2.
DR PROSITE; PS0106; PDZ; 2.
SQ SEQUENCE 403 AA; 45838 MW; AB24FB6E1D3144FD CRC64;

Query Match 18.4%; Score 128; DB 4; Length 403;
Best Local Similarity 40.0%; Pred. No. 0.00032;
Matches 30; Conservative 22; Mismatches 19; Indels 4; Gaps 3;

QY 39 GFTLKGLEGEPLISKVEGGKADTLSSKLQAGDEVVHINEVTLS-SRKEAVSLVK 97
DB 99 GLSVRGLEFGCGLFISHIKGGQADSVG--LQVGDEIVRINGYSISCTHEEVINLIR- 155

QY 98 SYKTLRLVVRNGVL 112
DB 156 TKKTVSIKVRHIGLI 170

RESULT 9
QY9S65 PRELIMINARY; PRT; 548 AA.
AC Q9S65;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE HARMONIN ISOFORM A1.
GN 2010016F01RIK OR USH1C.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20428180; PubMed=10973247;
RA Verpy E., Leibovici M., Zwaenepoel I., Liu X.Z., Gal A., Salem N.,
RA Mansour A., Blanchard S., Kobayashi I., Keats B.J., Slim R., Petit C.;
RT "A defect in harmonin, a PDZ domain-containing protein expressed in
RT the inner ear sensory hair cells, underlies Usher syndrome type 1C.";
RL Nat. Genet. 26:51-55(2000).
DR EMBL; AF228924; AAC12457.1; .
DR MGD; MGI:1919338; 2010016F01RIK.
DR InterPro; IPR001478; PDZ.
DR Pfam; PF00595; PDZ; 3.
DR SMART; SM00228; PDZ; 3.
DR PROSITE; PS0106; PDZ; 3.
SQ SEQUENCE 548 AA; 61712 MW; 44A245AA32091070 CRC64;

Query Match 18.4%; Score 128; DB 11; Length 548;
Best Local Similarity 40.0%; Pred. No. 0.00047;
Matches 30; Conservative 22; Mismatches 19; Indels 4; Gaps 3;

DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE ANTIGEN NY-CO-37.
GN NY-CO-38.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-COLON CANCER METASTASIS TO LUNG;
RX MEDLINE=9627252; PubMed=9610721;
RA Scanlan M.J., Chen Y.T., Williamson B., Gure A.O., Stockert E.,
RA Gordon J.D., Tureci O., Sahin U., Pfreundschuh M., Old L.J.;
RT "Characterization of human colon cancer antigens recognized by
RT autologous antibodies.";
RL Int. J. Cancer 76:652-658(1998).
DR EMBL; AF039699; AAC18048.1; .
DR HSSP; P29476; 1QAV.
DR InterPro; IPR001478; PDZ.
DR Pfam; PF00595; PDZ; 2.
DR SMART; SM00228; PDZ; 2.
DR PROSITE; PS0106; PDZ; 2.
SQ SEQUENCE 403 AA; 45838 MW; AB24FB6E1D3144FD CRC64;

Query Match 18.4%; Score 128; DB 4; Length 552;
Best Local Similarity 40.0%; Pred. No. 0.00047;
Matches 30; Conservative 22; Mismatches 19; Indels 4; Gaps 3;

QY 39 GFTLKGLEGEPLISKVEGGKADTLSSKLQAGDEVVHINEVTLS-SRKEAVSLVK 97
DB 99 GLSVRGLEFGCGLFISHIKGGQADSVG--LQVGDEIVRINGYSISCTHEEVINLIR- 155

QY 98 SYKTLRLVVRNGVL 112
DB 156 TKKTVSIKVRHIGLI 170

RESULT 11
QY9G69 PRELIMINARY; PRT; 652 AA.
AC Q9G69;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE ANTIGEN NY-CO-38.
GN NY-CO-38.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98272252; PubMed=9610721;
RA Scanlan M.J., Chen Y.T., Williamson B., Gure A.O., Stockert E.,
RA Gordon J.D., Tureci O., Sahin U., Pfreundschuh M., Old L.J.;
RT "Characterization of human colon cancer antigens recognized by
```

Best Local Primary 2007, 2008
Matches 29; Conservative 15; Mismatches 28; Indels 3; Gaps 3

```

DE AIE-75.
GN AIE-75.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=SMALL INTESTINE;
RX MEDLINE=9431904; PubMed=10500064;
RA Kobayashi I., Imamura K., Kubota M., Ishikawa S., Yamada M.,
RA Tonoki H., Okano M., Storch W.F., Morluchi T., Sakiyama Y.,
RT Kobayashi K.;
RA "Identification of an autoimmune enteropathy-related 75-kilodalton
RT antigen.";
RL Gastroenterology 117:823-830(1999).
DR EMBL; AB000955; BAA81739.1; -.
DR HSSP; P29476; IOAV.
DR InterPro; IPR001478; PD2.
DR SMART; PF00595; PDZ; 3.
DR SMART; SM00228; PDZ; 3.
DR PROSITE; PS0106; PDZ; 3.
SQ SEQUENCE 521 AA; 58304 MW; 168CCC58C33B9CCE CRC64;

Query Match 18.1%; Score 126; DB 4; Length 521;
Best Local Similarity 40.0%; Pred. No. 0.00068;
Matches 30; Conservative 21; Mismatches 20; Indels 4; Gaps 3;

QY 39 GFTLKGGLGEGEPLIISKVEGGKADTLSSKQAGDEVVHINEVTLS-SRKEAVSLVKG 97
DB 68 GLSVSGGLEFGCGLFISHLINGQADSFG--LQVGDEIVRINGYSISCTHEEVINLR- 124
QY 98 SYKTLRLVVRNGVL 112
DB 125 TKKTIVSIKVRHIGLI 139

Search completed: January 31, 2002, 21:05:35
Job time: 98 sec

```

```

DE AIE-75.
GN AIE-75.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=SMALL INTESTINE;
RX MEDLINE=9431904; PubMed=10500064;
RA Kobayashi I., Imamura K., Kubota M., Ishikawa S., Yamada M.,
RA Tonoki H., Okano M., Storch W.F., Morluchi T., Sakiyama Y.,
RT Kobayashi K.;
RA "Identification of an autoimmune enteropathy-related 75-kilodalton
RT antigen.";
RL Gastroenterology 117:823-830(1999).
DR EMBL; AB000955; BAA81739.1; -.
DR HSSP; P29476; IOAV.
DR InterPro; IPR001478; PD2.
DR SMART; PF00595; PDZ; 3.
DR SMART; SM00228; PDZ; 3.
DR PROSITE; PS0106; PDZ; 3.
SQ SEQUENCE 521 AA; 58304 MW; 168CCC58C33B9CCE CRC64;

Query Match 18.3%; Score 127.5; DB 11; Length 1171;
Best Local Similarity 27.3%; Pred. No. 0.0014;
Matches 39; Conservative 25; Mismatches 52; Indels 27; Gaps 6;

QY 9 HKPS--ATLNSNTATKGRY-----IYLEAFLEGGAPMGFTLKGGLGEGE 50
DB 998 HAPSQGTQETRTTKPKDQSOFEFKPQAQEQDYTYVELERGAKGFGFSLRGGRYNN 1057
QY 51 PLIISKVEGGKADTLSSKQAGDEVVHIN-EVTLSSSRKEAVSLVKGSKYKTLRLVVRN 109
DB 1058 DLVLRLAEDGPAER-CGKNRIGDEILEINGETTKNKKHSRAIELIKNGRRVRLFRRG 1116
QY 110 GVLLCCPEW-----KVVAIHRHD 127
DB 1117 DGSV--PEVAMIPPKIAACNRNE 1137

RESULT 15
Q9UM17 PRELIMINARY; PRT; 521 AA.
AC Q9UM17
DT 01-JUN-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)

```

Fri Feb - 1 08:56:37 2002

us-09-641-831-6.rsp

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OK protein - protein search, using sw model

Run on: January 31, 2002, 21:06:28 ; Search time 26.99 Seconds
(without alignments)
146.714 Million cell updates/sec

US-09-641-831-4

Title: 545
Perfect score: 1 MMRTDFHFKPSATLNSNTA.....KEAVSLVKGSKTLRLVYRS 108
Sequence:

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 08
Maximum Match 1008

Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	270.5	49.6	1616	1 APXL_HUMAN	Q13796 homo sapien
2	126	23.1	852	1 DLG2_RAT	Q63622 rattus norv
3	126	23.1	870	1 DLG3_HUMAN	Q15700 homo sapien
4	122.5	22.5	913	1 PTN3_HUMAN	P26045 homo sapien
5	113.5	20.8	327	1 PDL1_RAT	P52944 rattus norv
6	112.5	20.6	330	1 RIL_RAT	P36202 rattus norv
7	112.5	20.6	911	1 DLG1_RAT	Q62696 rattus norv
8	111	20.4	904	1 DLG1_HUMAN	Q12959 homo sapien
9	108.5	19.9	330	1 RIL_CHICK	Q9PW72 gallus gall
10	106.5	19.5	849	1 DLG3_MOUSE	P70175 mus musculu
11	106.5	19.5	849	1 DLG3_RAT	Q62936 rattus norv
12	104.5	19.2	329	1 PDL1_HUMAN	Q00151 homo sapien
13	104.5	19.2	724	1 DLG4_MOUSE	Q62108 mus musculu
14	104.5	19.2	724	1 DLG4_RAT	P31016 rattus norv
15	104.5	19.2	724	1 DLG4_HUMAN	P78332 homo sapien
16	99.5	18.3	817	1 DLG3_HUMAN	Q92766 homo sapien
17	99	18.2	1026	1 PTN1_CAEEL	P28191 caenorhabd1
18	98.5	18.1	926	1 PTN4_HUMAN	P29074 homo sapien
19	98.5	18.1	960	1 DLG1_DROME	P31007 drosophila
20	98.5	18.1	1387	1 RGSC_RAT	Q08774 rattus norv
21	96	17.6	326	1 PDL1_MOUSE	Q70400 mus musculu
22	95	17.4	328	1 RIL_HUMAN	P50479 homo sapien
23	92.5	17.0	1447	1 RGSC_HUMAN	Q14924 homo sapien
24	83.5	15.3	898	1 Z03_CANFA	Q62663 canis fam11
25	83.5	15.3	933	1 Z03_HUMAN	Q95049 homo sapien
26	83	15.2	643	1 RHOP_MOUSE	Q61085 mus musculu
27	82.5	15.1	2485	1 PTND_HUMAN	Q12923 homo sapien
28	80	14.7	385	1 MPB3_HUMAN	Q13368 homo sapien
29	79.5	14.6	905	1 Z03_MOUSE	Q9QXY1 mus musculu
30	78	14.3	631	1 IL16_HUMAN	Q14005 homo sapien
31	73.5	13.5	3988	1 POLG_BVDVN	P19711 bovine vira
32	72.5	13.3	3898	1 POLG_BVDVN	Q01499 bovine vira
33	72	13.2	145	1 OM25_HUMAN	P57105 homo sapien

34	72	13.2	1190	1 Z02_HUMAN	Q9UDY2 homo sapien
35	72	13.2	1736	1 Z01_HUMAN	Q07157 homo sapien
36	72	13.2	1745	1 Z01_MOUSE	P39447 mus musculu
37	71.5	13.1	3898	1 POLG_HCVB	P19712 hog cholera
38	71.5	13.1	3898	1 POLG_HCVB	P21530 hog cholera
39	71	13.0	322	1 CYS1_HUMAN	P13277 homarus ame
40	71	13.0	466	1 EM55_MOUSE	P70290 mus musculu
41	71	13.0	1167	1 Z02_MOUSE	Q920U1 mus musculu
42	71	13.0	1174	1 Z02_CANFA	Q95168 canis fam11
43	70	12.8	206	1 OM25_RAT	Q9VWJ4 rattus norv
44	69.5	12.8	1427	1 ABC1_SCHPO	Q92337 schizosacch
45	69	12.7	1591	1 T1AM_HUMAN	Q13009 homo sapien

ALIGNMENTS

RESULT	ID	APXL_HUMAN	STANDARD:	PRT:	1616 AA.
AC	Q13796	01-NOV-1997 (rel. 35, Created)			
DT	01-NOV-1997 (rel. 35, Last sequence update)				
DT	01-NOV-1997 (rel. 35, Last sequence update)				
DT	20-AUG-2001 (rel. 40, Last annotation update)				
DE	APICAL-LIKE PROTEIN (APXL PROTEIN).				
GN	APXL				
OS	Homo sapiens (Human)				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.				
ON	NCBI_TaxID=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=Retina;				
RX	MEDLINE=95315933; PubMed=7795590;				
RA	Schiaffino V.M., Bassi M.T., Ruggeri E.I., Renieri A., Galli L.,				
RA	Ballabio A.;				
RT	"Cloning of a human homologue of the Xenopus laevis Apx gene from the				
RT	ocular albinism type 1 critical region.";				
RL	Hum. Mol. Genet. 4:373-382(1995).				
RN	[2]				
RP	SEQUENCE OF 56-1616 FROM N.A.				
RA	Shen Y., Glabbs R.A.;				
RL	Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.				
CC	- TISSUE SPECIFICITY: ABUNDANT IN RETINA AND MELANOMA; ALSO IN				
CC	BRAIN, PLACENTA, LUNG, KIDNEY, PANCREAS.				
CC	- SIMILARITY: SOME, TO XENOPUS LAEVIS APICAL PROTEIN (APX).				
CC	- This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@sib-sib.ch).				
CC	-----				
DR	EMBL: X83543; CAAS8534.1; -				
DR	EMBL: AC002365; AAC32592.1; -				
DR	HSSP: Q12959; IPDR.				
DR	MIM: 300103; -				
DR	InterPro: IPR000867; IGFBP.				
DR	InterPro: IPR001478; PDZ.				
DR	Pfam: PF00219; IGFBP; 1.				
DR	Pfam: PF00595; PDZ; 1.				
DR	SMART: SM00228; PDZ; 1.				
DR	PROSITE: PS50106; PDZ; 1.				
FT	SMILAR	54	112		
FT					
FT	DOMAIN	150	153		
FT	DOMAIN	314	320		
FT	DOMAIN	343	346		
FT	DOMAIN	1065	1068		
FT	DOMAIN	1616 AA;	176409 MW;		
FT	SEQUENCE	1616 AA;	176409 MW;		

TO MOUSE SYNTROPHIN-1 AND RABBIT
DYSTROPHIN-ASSOCIATED PROTEIN.

Fri Feb 1 08:56:32 2002

Query Match 49.6%; Score 270.5; DB 1; Length 1616;
Best Local Similarity 58.9%; Pred. No. 3.3e-19;
Matches 56; Conservative 14; Mismatches 22; Indels 3; Gaps 2;

QY 13 ATLSNTATKGRVYLEAFLEGGAPWGFTLKGGLEHGEPLIIISKVEEGKADTLSSKLOA 72
DB 14 AEAETRADGGRLV--EVQLSGAPWGFTLKGGREHGEPLIVTKIEGSKAAAV-DKLLA 70

QY 73 GDEVVHINEVTLSSSRKENVSLVSGSYKTLRLVLR 107
DB 71 GDEIVGINDIGLSGPRQPAICLVKSGHKTLLVVK 105

RESULT 2
DLG2_RAT STANDARD; PRT; 852 AA.
ID DLG2_RAT Q63622; Q62939; P70548;
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE CHANNEL ASSOCIATED PROTEIN OF SYNAPSE-110 (CHAPSYN-110) (SYNAPTIC CHANNEL ASSOCIATED PROTEIN PSD-93) (DISCS, LARGE HOMOLOG 2).
GN DLG2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
[1]
RN SEQUENCE FROM N.A.
RA MEDLINE=96310881; PubMed=8755482;
RX Kim E., Cho K.O., Rothschild A., Sheng M.;
RT "Heteromultimerization and NMDA receptor-clustering activity of ChapSyn-110, a member of the PSD-95 family of proteins.";
RL Neuron 17:103-113(1996).
[2]
RN SEQUENCE FROM N.A.
RA MEDLINE=96193770; PubMed=8625413;
RX Brennan J.E., Chao D.S., Gee S.H., McGee A.W., Craven S.E., Santillano D.R., Wu Z., Huang F., Xia H., Peters M.F., Froehner S.C., Bredt D.S.;
RT "Interaction of nitric oxide synthase with the postsynaptic density protein PSD-95 and alpha1-syntrophin mediated by PDZ domains.";
RL Cell 84:757-767(1996).
[3]
RN SEQUENCE FROM N.A.
RA Irie M., Hata Y., Takai Y.;
RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: INTERACTS WITH THE CYTOPLASMIC TAIL OF NMDA RECEPTOR SUBUNITS AS WELL AS POTASSIUM CHANNELS.
CC -1- SIMILARITY: CONTAINS 3 PDZ/DHR DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -1- SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U49049; AAB53243.1; .
DR EMBL: U50717; AAC52643.1; .
DR EMBL: U53368; AAB48562.1; .
DR HSP: Q12959; IPDR.
DR InterPro: IPR000619; Guanylate_kin.
DR InterPro: IPR001478; PDZ.
DR InterPro: IPR001452; SH3.
DR Pfam: PF00625; Guanylate_kin; 1.
DR Pfam: PF00595; PDZ; 3.
DR Pfam: PF00018; SH3; 1.

SMART; SM00072; GuKc; 1.
SMART; SM00228; PDZ; 3.
SMART; SM00326; SH3; 1.
PROSITE; PS00856; GUANYLATE_KINASE_1; 1.
PROSITE; PS00552; GUANYLATE_KINASE_2; 1.
PROSITE; PS0106; PDZ; 3.
PROSITE; PS00002; SH3; 1.
KW SH3 domain; Repeat.
FT DOMAIN 98 184
FT DOMAIN 193 279
FT DOMAIN 421 501
FT DOMAIN 536 606
FT DOMAIN 662 852
FT DOMAIN 181 182
FT CONFLICT 228 228
FT CONFLICT 326 326
FT CONFLICT 339 339
FT CONFLICT 450 454
FT CONFLICT 464 465
FT CONFLICT 474 474
FT CONFLICT 476 476
FT CONFLICT 478 478
FT CONFLICT 484 486
FT CONFLICT 506 506
FT CONFLICT 569 569
FT CONFLICT 586 586
FT CONFLICT 626 641
FT CONFLICT 639 639
FT CONFLICT 726 726
FT CONFLICT 733 733
FT CONFLICT 749 749
FT CONFLICT 756 756
FT CONFLICT 791 792
FT CONFLICT 794 794
SQ SEQUENCE 852 AA; 94934 MW; F8D414A8B9CF5B09 CRC64;

Query Match. 23.1%; Score 126; DB 1; Length 852;
Best Local Similarity 32.7%; Pred. No. 5.3e-05;
Matches 33; Conservative 26; Mismatches 34; Indels 8; Gaps 5;

QY 10 KPSATLSNATKQ--RYVILEAPLEGAPWGFTLKGGLEHGEPLIIISKVEEGKADTLSS 67
DB 404 QPSVTLQRAISLEGEPRKVVLA---KGTGGLGFNIVG--EDGEGIFVSFILAGGPAD-LS 458

QY 68 SKLAQGEVHVHINEVTL-SSSRKENVSLVSGSYKTLRLVLR 107
DB 459 GELQRGDQILSVNGIDLRGASHEQAALAKGAGQTVTIIAQ 499

RESULT 3
DLG2_HUMAN STANDARD; PRT; 870 AA.
ID DLG2_HUMAN Q15700;
AC Q15700;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE CHANNEL ASSOCIATED PROTEIN OF SYNAPSE-110 (CHAPSYN-110) (DISCS, LARGE HOMOLOG 2).
GN DLG2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
RN SEQUENCE FROM N.A.
RA MEDLINE=96310881; PubMed=8755482;
RX Kim E., Cho K.O., Rothschild A., Sheng M.;
RT "Heteromultimerization and NMDA receptor-clustering activity of ChapSyn-110, a member of the PSD-95 family of proteins.";
RL Neuron 17:103-113(1996).
CC -1- FUNCTION: INTERACTS WITH THE CYTOPLASMIC TAIL OF NMDA RECEPTOR SUBUNITS AS WELL AS POTASSIUM CHANNELS.

Db 479 DFHR--VTGGSTASQYDKNDGSDYLVLRITPDEDGKFGNKLKGGVDQKMPLV 536
QY 54 ISKVEEGKADTLSSKLOAGDEVVHNEVTLSS-SRKEAVSLVKGYSKT-----LRLVVR 107
Db 537 VSRINPSPADTCIPKLNQGDQIVLNGRDISEHTHQVVMFKASRSHRELALVIR 595

RESULT 5
ID PDLL_RAT STANDARD; PRT; 327 AA.
AC P52944;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PD2 AND LIM DOMAIN PROTEIN 1 (LIM DOMAIN PROTEIN CLP-36) (C-TERMINAL
DE LIM DOMAIN PROTEIN 1) (ELFIN).
GN PDLM1 OR CLIM1 OR CLP36.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SPRAGUE-DAWLEY; TISSUE=Liver;
RX MEDLINE=96096533; PubMed=8522188;
RA Wang H., Harrison-Shostak D.C., Lemasters J.J., Herman B.;
RT "Cloning of a rat cDNA encoding a novel LIM domain protein with high
RL homology to rat RIL".
RL Gene 165:267-271(1995).
CC -!- FUNCTION: CYTOSKELETAL PROTEIN THAT MAY ACT AS AN ADAPTER THAT
CC BRINGS OTHER PROTEINS TO THE CYTOSKELETON (BY SIMILARITY).
CC -!- SUBUNIT: INTERACTS WITH ALPHA-ACTININS, 1 2 AND 4 (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: CYTOSOL; CYTOSOL (BY SIMILARITY).
CC -!- TISSUE SPECIFICITY: EXPRESSED MOST ABUNDANTLY IN HEART, LUNG AND
CC LIVER, MODERATELY IN SPLEEN AND SKELETAL MUSCLE, AND AT EXTREMELY
CC LOW LEVELS (IF AT ALL) IN TESTIS AND BRAIN TISSUES.
CC -!- SIMILARITY: CONTAINS 1 LIM DOMAIN. THE LIM DOMAIN BINDS 2 ZINC
CC IONS.
CC -!- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U23769; AAA92046.1; -
CC InterPro; IPR001781; LIM.
CC InterPro; IPR001478; PDZ.
CC Pfam; PF00412; LIM; 1.
CC Pfam; PF00595; PDZ; 1.
CC ProDom; PD000094; LIM; 1.
CC SMART; SM00132; LIM; 1.
CC SMART; SM00228; PDZ; 1.
CC PROSITE; PS00478; LIM_DOMAIN_1; 1.
CC PROSITE; PS50023; LIM_DOMAIN_2; 1.
CC PROSITE; PS50106; PDZ; 1.
CC Cytoskeleton; LIM domain; Metal-binding; Zinc.
KW Cytoskeleton; LIM domain; Metal-binding; Zinc.
FT DOMAIN 3 84 PDZ.
FT DOMAIN 258 308 LIM.
SQ SEQUENCE 327 AA; 35525 MW; 4C3B5F579FBDA4D1 CRC64;

Query Match 20.8%; Score 113.5; DB 1; Length 327;
Best Local Similarity 39.5%; Pred. No. 0.00032;
Matches 30; Conservative 11; Mismatches 32; Indels 3; Gaps 2;
QY 32 LEGGAPWGFTLKGLEHCEPLIISKVEEGKADTLSSKLOAGDEVVHIN-EVTLSSSRKE 90
Db 8 LOGPGPWGFRVLGGKDFEQLAISRTVPGSKA--AIANLCIGDLTAIDGDTSSMTLHE 65

QY 91 AVSLVKGYSKTLRLVV 106
Db 66 AQNRKKGCVDMNTLV 81

RESULT 6
ID RIL_RAT STANDARD; PRT; 330 AA.
AC P36202;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE LIM PROTEIN RIL (REVERSION-INDUCED LIM PROTEIN) (RIT-18) (H-REV18).
GN RIL.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SPRAGUE-DAWLEY; TISSUE=Liver;
RX MEDLINE=95124724; PubMed=7824279;
RA Kiess M., Scharm B., Aguzzi A., Hajnal A., Klemeniz R.,
RA Schwarte-Waldhoff I., Schaefer R.;
RT "Expression of ril, a novel LIM domain gene, is down-regulated in
RT Hras-transformed cells and restored in phenotypic revertants".
RL Oncogene 10:61-68(1995).
CC -!- TISSUE SPECIFICITY: DETECTED IN SEVERAL TISSUES, MOST PROMINENT IN
CC BRAIN AND HEART OF ADULTS.
CC -!- SIMILARITY: CONTAINS 1 LIM DOMAIN. THE LIM DOMAIN BINDS 2 ZINC
CC IONS.
CC -!- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X76454; CAA53992.1; -
CC PIR; S39246; S39246.
CC InterPro; IPR001781; LIM.
CC InterPro; IPR001478; PDZ.
CC Pfam; PF00412; LIM; 1.
CC Pfam; PF00595; PDZ; 1.
CC ProDom; PD000094; LIM; 1.
CC SMART; SM00132; LIM; 1.
CC SMART; SM00228; PDZ; 1.
CC PROSITE; PS00478; LIM_DOMAIN_1; 1.
CC PROSITE; PS50023; LIM_DOMAIN_2; 1.
CC PROSITE; PS50106; PDZ; 1.
CC LIM domain; Metal-binding; Zinc.
KW LIM domain; Metal-binding; Zinc.
FT DOMAIN 1 84 PDZ.
FT DOMAIN 255 305 LIM.
SQ SEQUENCE 330 AA; 35521 MW; 84BEB1E629ADA584 CRC64;

Query Match 20.6%; Score 112.5; DB 1; Length 330;
Best Local Similarity 40.8%; Pred. No. 0.00041;
Matches 31; Conservative 9; Mismatches 33; Indels 3; Gaps 2;

QY 32 LEGGAPWGFTLKGLEHCEPLIISKVEEGKADTLSSKLOAGDEVVHIN-EVTLSSSRKE 90
Db 7 LRGPSPWGFRLVGGKDFEQLAISRTVPGSKA--ALAAALCPGDSIQANGESTLMTLHE 64
QY 91 AVSLVKGYSKTLRLVV 106
Db 65 AQNRKKGCHDLTLV 80

ID	DUGL_RAT	STANDARD	PRT	911 AA.
RESULT 7				
DUGL_RAT				
AC	062696			
DT	01-NOV-1997 (Rel. 35, Created)			
DT	01-NOV-1997 (Rel. 35, Last sequence update)			
DT	20-AUG-2001 (Rel. 40, Last annotation update)			
DE	PRESYNAPTIC PROTEIN SAP97 (SYNAPSE-ASSOCIATED PROTEIN 97) (DISCS, LARGE HOMOLOG 1).			
DE	LARGE HOMOLOG 1).			
GN	DUGL			
OC	Rattus norvegicus (Rat).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrate; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.			
OX	NCBI_TaxID=10116;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Brain;			
RX	MEDLINE=95198112; PubMed=7891172;			
RA	Mueller B.M., Kistner U., Voh R.W., Cases-Langhoff C., Becker B.,			
RA	Gundelfinger E.D., Garner C.C.;			
RT	"Molecular characterization and spatial distribution of SAP97, a novel presynaptic protein homologous to SAP90 and the Drosophila discs-large tumor suppressor protein."			
RL	J. Neurosci. 15:2354-2366(1995).			
CC	-1- FUNCTION: INTERACTS WITH THE CYTOPLASMIC TAIL OF NMDA RECEPTOR SUBUNITS. MAY BE INVOLVED IN SYNAPTOGENESIS. MAY PLAY A ROLE IN CELL ADHESION.			
CC	-1- SUBCELLULAR LOCATION: IN CULTURED T84 CELLS, IT IS RESTRICTED TO THE CYTOPLASMIC SURFACE OF THE PLASMA MEMBRANE BETWEEN ADJACENT CELLS, BUT NOT AT THE EDGES OF CELLS LACKING CELL-CELL CONTACT.			
CC	-1- TISSUE SPECIFICITY: PRESYNAPTIC NERVE TERMINI OF EXCITATORY SYNAPSES. IN OTHER BRAIN REGIONS, SAP97 IS FOUND IN AND ALONG BUNDLES OF UNMYELINATED AXONS. SAP97 IS NOT RESTRICTED TO THE CNS, BUT IS ALSO PRESENT AT THE BASAL LATERAL MEMBRANE BETWEEN A VARIETY OF EPITHELIAL CELLS.			
CC	-1- SIMILARITY: CONTAINS 3 PDZ/DHR DOMAINS.			
CC	-1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.			
CC	-1- SIMILARITY: CONTAINS A GUANYLATE KINASE-LIKE DOMAIN.			
CC	-1- SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sdb.ch/announce/ or send an email to license@isb-sdb.ch).			
CC	EMBL: U14950; AAA79976.1; -			
DR	HSSP: Q12959; 1PDR.			
DR	InterPro: IPR000619; Guanylate_kin.			
DR	InterPro: IPR001478; PDZ.			
DR	InterPro: IPR001452; SH3.			
DR	pfam: PFO0625; Guanylate_kin; 1.			
DR	pfam: PFO00595; PDZ; 3.			
DR	pfam: PFO0018; SH3; 1.			
DR	SMART: SM00072; GUKC; 1.			
DR	SMART: SM00228; PDZ; 3.			
DR	SMART: SM00326; SH3; 1.			
DR	PROSITE: PS00856; GUANYLATE_KINASE_1; 1.			
DR	PROSITE: PS00052; GUANYLATE_KINASE_2; 1.			
DR	PROSITE: PS50106; PDZ; 3.			
DR	PROSITE: PS50002; SH3; 1.			
KW	SH3 domain; Repeat.			
FT	DOMAIN 224 310 PDZ 1.			
FT	DOMAIN 318 404 PDZ 2.			
FT	DOMAIN 465 545 PDZ 3.			
FT	DOMAIN 580 650 SH3.			
FT	DOMAIN 721 911 GUANYLATE_KINASE.			
FT	DOMAIN 527 530 POLY-ALA.			
SEQUENCE	911 AA; 100570 MW; 18CEB31DD0CAFB8 CRC64;			

```

Query Match      20.6%; Score 112.5; DB 1; Length 911;
Best Local Similarity 27.9%; Pred. No 0.0013;
Matches 34; Conservative 27; Mismatches 42; Indels 19; Gaps 5;

OY      3 RTTEFHKPSATLMSNMTATKGRVYL-EARL-----BGAPMGFTLKGCL 46
Db      OSVNHSSPSVYLDGTPASPARYSPISKAVLGDDEITREPRKVVILHSGSTGLGFIIVGG- 482
OY      47 EHEGPLITSKVEEGKADTLSSKIQAGDEVHINEVTL-SSSRKENAVSYKGVSKTLRLV 105
Db      483 EDGEIIFISFTIAGGPAD-LSGELRKRGRIISVNSVDLRAASHEDQAAAALNAGCAVTV 541
OY      106 VR 107
Db      542 AQ 543

RESULT      8
DIG1_HUMAN STANDARD; PRT; 904 AA.
AC      012959; 012958;
DT      01-NOV-1997 (Rel. 35, Created)
DT      01-NOV-1997 (Rel. 35, Last sequence update)
DT      20-AUG-2001 (Rel. 40, Last annotation update)
DE      PRESYNAPTIC PROTEIN SAP97 (SYNAPSE-ASSOCIATED PROTEIN 97) (DISCS,
DE      LARGE HOMOLOG 1).
OS      DIG1.
GN      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX      NCBI_TaxID=9606;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=95024052; PubMed=7937897;
RA      Iue R.A., Marfatia S.M., Branton D., Chishti A.H.;
RT      "Cloning and characterization of hd19: the human homologue of the
RT      Drosophila discs large tumor suppressor binds to protein 4.1.",
RT      Proc. Natl. Acad. Sci. U.S.A. 91:9818-9822(1994).
RN      [2]
RN      X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS) OF 460-555.
RX      MEDLINE=96338231; PubMed=8757139;
RA      Cabral J.H., Petosa C., Sutcliffe M.J., Raza S., Byron O., Poy F.,
RA      Marfatia S.M., Chishti A.H., Liddington R.C.;
RT      "Crystal structure of a PDZ domain.",
RT      Nature 382:649-652(1996).
CC      -1- FUNCTION: INTERACTS WITH THE CYTOPLASMIC TAIL OF NMDA RECEPTOR
CC      SUBUNITS. ASSOCIATES WITH PROTEIN 4.1.
CC      -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: 1 (SHOWN HERE) AND 2; ARE
CC      PRODUCED BY ALTERNATIVE SPLICING.
CC      -1- SIMILARITY: CONTAINS 3 PDZ/DRH DOMAINS.
CC      -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC      -1- SIMILARITY: CONTAINS A GUANYLATE KINASE-LIKE DOMAIN.
CC      -1- SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS.
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation
CC      at the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL: U13897; AAA50599.1; -
DR      EMBL: U13896; AAA50598.1; -
DR      PDB: 1PDR; 23-JUL-97.
DR      MIM: 601014; -
DR      InterPro: IPR000619; Guanylate_kin.
DR      InterPro: IPR001478; PDZ.
DR      InterPro: IPR001452; SH3.
DR      Pfam: PF00625; Guanylate_kin; 1.
DR      Pfam: PF00595; PDZ; 3.
DR      Pfam: PF00018; SH3; 1.
DR      SMART: SM00072; GUKC; 1.

```

Fri Feb 1 08:56:32 2002

```

DR SMART; SMO00228; PDZ; 3.
DR SMART; SMO0326; SH3; 1.
DR PROSITE; PS00856; GUANYLATE_KINASE_1; 1.
DR PROSITE; PS00852; GUANYLATE_KINASE_2; 1.
DR PROSITE; PS50106; PDZ; 3.
DR PROSITE; PS50002; SH3; 1.
DR SH3 domain; Repeat: Alternative splicing; 3D-structure.
KW DOMAIN 224 310 PDZ 1.
FT FT DOMAIN 319 405 PDZ 2.
FT FT DOMAIN 466 546 PDZ 3.
FT FT DOMAIN 581 651 SH3.
FT FT DOMAIN 714 904 GUANYLATE_KINASE.
FT VARSPLIC 669 680
FT SEQUENCE 904 AA; 100354 MW; B78798D6BB0920D4 CRC64;
SQ

Query Match 20.4%; Score 111; DB 1; Length 904;
Best Local Similarity 30.3%; Pred. No. 0.0018;
Matches 36; Conservative 23; Mismatches 40; Indels 20; Gaps 6;

OY 7 DFH-RPSATLNSNTATKGRVIYL-EAPL-----EGCAPGFTLKGGLG 49
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 428 DNHVSFFSLGQTSPASRPVSPVKVSGVLGDDETRPVKVLRHSGTGLGFNIVGG-EDG 486
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 50 EPLIITSKYVEEGKADTLSSKIQAGDEVVHINEVTL-SSSRKEAVSLVKGSYKTLRLVWR 107
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 487 EGIFTSFILAGPAD-LSGELRKGDRIISVSDLRASHEQAAALKNAGQAVTIVAQ 544
| | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 9
RIL_CHICK
ID RIL_CHICK STANDARD; PRT; 330 AA.
AC Q9PW72.
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE LIM PROTEIN RIL (REVERSION-INDUCED LIM PROTEIN).
DE RIL.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
OC Gallus.
OC NCBI_TaxID=9031;
OX [1]
RN R1.
RP SEQUENCE FROM N.A.
RC STRAIN-White leghorn;
RX MEDLINE=20379359; PubMed=10918612;
FA Fu S.-L., Waha A., Vogt P.K.;
RT "Identification and characterization of genes upregulated in cells
transformed by v-Jun.";
RL Oncogene 19:3537-3545(2000).
CC -1- FUNCTION: NEGATIVE REGULATOR OF CELL GROWTH.
CC -1- SIMILARITY: CONTAINS 1 LIM DOMAIN. THE LIM DOMAIN BINDS 2 ZINC
CC IONS.
CC -1- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercia
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement
CC or send an email to license@isb-sib.ch.)
CC
CC EMBL; AF167295; AAD46655.1; -.
CC InterPro; IPR001781; LIM.
CC InterPro; IPR001478; PDZ.
CC Pfam; PF00412; LIM; 1.
CC Pfam; PF00595; PDZ; 1.
CC ProDom; PD000094; LIM; 1.
CC SMART; SM00132; LIM; 1.
CC SMART; SMO0228; PDZ; 1.

```

FT DOMAIN 149 235 PDZ 1.
FT DOMAIN 244 330 PDZ 2.
FT DOMAIN 404 484 PDZ 3.
FT DOMAIN 519 589 SH3.
FT DOMAIN 659 849 GUANYLATE KINASE.
SO SEQUENCE 849 AA; 93482 MM; EF3EF2D/51538EE CRC64;

Query Match 19.5%; Score 106.5; DB 1; Length 849;
Best Local Similarity 32.4%; Pred. No. 0.0048; Indels 11; Gaps 6;
Matches 34; Conservative 23; Mismatches 37;

QY 10 KPSATLNSNTKRGRIYILEAFLE-GGAPMGFTLKGLEH---GEP-LIISKVEEGKA 63
DB 133 EPSLSVN---GSDGMFKYEELVLEKNGSLGFSIAGGIDNHPVDDPGIFTKIIPGGA 189
DB 64 DTLSSKIQAGDEVVHINEVTLSS-SRKEAVSLVSGSYKTLVLVR 107
DB 190 -AMGRLGVNDCVLRVNEVDSEVHSHRAVEALKRAGVRLVVR 233

RESULT 11
ID DILG3_RAT STANDARD; PRT; 849 AA.

AC 062936; P70547; (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PRESYNAPTIC PROTEIN SAP102 (SYNAPSE-ASSOCIATED PROTEIN 102) (PSD-95/SAP90 RELATED PROTEIN 1) (DISCS, LARGE HOMOLOG 3).
GN DILG3.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_Taxid-10116;
RN [1]
RP SEQUENCE FROM N.A. (LONG FORM).
RC TISSUE-Brain;
RX MEDLINE-96374358; PubMed-8780649;
RA Mueller B.M., Kistner U., Kindler S., Chung M.J., Kuhlendahl S., Penster S.D., Lau L.-F., Voh R.W., Huganir R.L., Gundelfinger E.D., Garner C.C.;
RA "SAP102, a novel postsynaptic protein that interacts with NMDA receptor complexes in vivo.";
RA Neuron 17:255-265(1996).
RN [2]
RP SEQUENCE FROM N.A. (SHORT FORM).
RA Title M., Hata Y., Takai Y.;
RA Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
RL -1- FUNCTION: INTERACTS WITH THE CYTOPLASMIC TAIL OF THE NMDA RECEPTOR SUBUNIT NR2B.
CC -1- SIMILARITY: CONTAINS 3 PDZ/DHR DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -1- SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL: U50147; AAA93031.1; -
DR EMBL: U53367; AAA48561.1; -
DR HSSP: Q12955; 1PDR.
DR InterPro: IPR000619; Guanylate_kin.
DR InterPro: IPR001478; PDZ.
DR InterPro: IPR001452; SH3.
DR Pfam: PF00625; Guanylate_kin; 1.
DR Pfam: PF00595; PDZ; 3.
DR Pfam: PF00018; SH3; 1.

DR SMART: SM00072; GUKC; 1.
DR SMART: SM00228; PDZ; 3.
DR SMART: SM00326; SH3; 1.
DR PROSITE: PS00856; GUANYLATE KINASE_1; 1.
DR PROSITE: PS50052; GUANYLATE_KINASE_2; 1.
DR PROSITE: PS50106; PDZ; 3.
DR PROSITE: PS50002; SH3; 1.
KM SH3 domain; Repeat; Alternative splicing.
FT DOMAIN 149 235 PDZ 1.
FT DOMAIN 244 330 PDZ 2.
FT DOMAIN 404 484 PDZ 3.
FT DOMAIN 519 589 SH3.
FT DOMAIN 659 849 GUANYLATE KINASE.
FT VARSPLIC MISSING (IN SHORT ISOFORM).
SO SEQUENCE 849 AA; 93539 MM; 34DA9C46C7BB96DB CRC64;

Query Match 19.5%; Score 106.5; DB 1; Length 849;
Best Local Similarity 32.4%; Pred. No. 0.0048; Indels 11; Gaps 6;
Matches 34; Conservative 23; Mismatches 37;

QY 10 KPSATLNSNTKRGRIYILEAFLE-GGAPMGFTLKGLEH---GEP-LIISKVEEGKA 63
DB 133 EPSLSVN---GSDGMFKYEELVLEKNGSLGFSIAGGIDNHPVDDPGIFTKIIPGGA 189
DB 64 DTLSSKIQAGDEVVHINEVTLSS-SRKEAVSLVSGSYKTLVLVR 107
DB 190 -AMGRLGVNDCVLRVNEVDSEVHSHRAVEALKRAGVRLVVR 233

RESULT 12
ID PDL1_HUMAN STANDARD; PRT; 329 AA.

AC 000151; 2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PDZ AND LIM DOMAIN PROTEIN 1 (LIM DOMAIN PROTEIN CLP-36) (C-TERMINAL LIM DOMAIN PROTEIN 1) (ELFIN).
GN PDLIM1 OR CLIM1 OR CLP36.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid-9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-Heart.
RX MEDLINE-99145108; PubMed-10022510;
RA Kotaka M., Ngai S.W., Garcia-Barcelo M., Tsui S.K.W., Fung K.P., Lee C.Y., Waye M.M.Y.;
RA "Characterization of the human 36-kDa carboxyl terminal LIM domain protein (hclim1).";
RA J. Cell. Biochem. 72:279-285(1999).
RN [2]
RP INTERACTION WITH ALPHA-ACTININ 2.
RX MEDLINE-20349365; PubMed-10861853;
RA Kotaka M., Kostin S., Ngai S., Chan K., Lau Y., Lee S.M., Li H.Y., Ng E.K., Schaper J., Tsui S.K.W., Fung K.P., Lee C.Y., Waye M.M.Y.;
RA "Interaction of hclim1, an enigma family protein, with alpha-actinin 2.";
RA J. Cell. Biochem. 78:558-565(2000).
RN [3]
RP INTERACTION WITH ALPHA-ACTININS 1 AND 4.
RX MEDLINE-20219155; PubMed-10753915;
RA Vallianus T., Luukko K., Makela T.P.;
RA "CLP-36 PDZ-LIM protein associates with nonmuscle alpha-actinin-1 and alpha-actinin-4.";
RT J. Biol. Chem. 275:11100-11105(2000).
CC -1- FUNCTION: CYTOSKELETAL PROTEIN THAT MAY ACT AS AN ADAPTER THAT BRINGS OTHER PROTEINS TO THE CYTOSKELETON.
CC -1- SUBUNIT: INTERACTS WITH ALPHA-ACTININS, 1 2 AND 4.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- TISSUE SPECIFICITY: STRONGLY EXPRESSED IN THE HEART AND SKELETAL

CC -1- TISSUE SPECIFICITY: STRONGLY EXPRESSED IN THE HEART AND SKELETAL

RX MEDLINE 93186749; PubMed=76803433; Cases-Langhoff C., Garner A.M.,
 RA Kistner U., Wenzel B.M., Voh R.W., Cases-Langhoff E.D., Garner C.C.;
 RA Appeltrauer U., Voss B., Gundelfinger E.D., Garner C.C.;
 RT "SAP90, a rat presynaptic protein related to the product of the
 RL Drosophila tumor suppressor gene d19-A.";
 RL J. Biol. Chem. 268:4580-4583(1993).
 RN [3]
 RN SEQUENCE OF 566-625 FROM N.A.
 RP STRAIN-WISTAR KYOTO; TISSUE=Vascular smooth muscle;
 RC Adams L.D., Werny I., Schwartz S.M.;
 RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RN X-RAY CRYSTALLOGRAPHY (1.82 ANGSTROMS) OF 302-402.
 RX MEDLINE-96270509; PubMed-8674113;
 RA Doyle D.A., Lee A., Lewis J., Kim E., Sheng M., Mackinnon R.;
 RT "Crystal structures of a complexed and peptide-free membrane protein-
 RL binding domain: molecular basis of peptide recognition by PDZ.";
 RL Cell 85:1067-1076(1996).
 CC -1- FUNCTION: INTERACTS WITH THE CYTOPLASMIC TAIL OF NMDA RECEPTOR
 CC SUBUNITS. MAY BE INVOLVED IN SYNAPTOGENESIS.
 CC -1- SUBCELLULAR LOCATION: CONCENTRATED AT SYNAPTIC JUNCTIONS PRIMARILY
 CC ON THE PRESYNAPTIC SIDE (WAS ORIGINALLY THOUGHT TO BE
 CC POSTSYNAPTIC).
 CC -1- TISSUE SPECIFICITY: PRESYNAPTIC DENSITY FRACTION OF BRAIN.
 CC -1- SIMILARITY: CONTAINS 3 PDZ/DRH DOMAINS.
 CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
 CC -1- SIMILARITY: CONTAINS A GUANYLATE KINASE-LIKE DOMAIN.
 CC -1- SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-slb.ch/announce/>
 CC or send an email to license@isb-slb.ch).
 CC -----
 DR EMBL; M96853; AAA41971.1; -;
 DR EMBL; X66474; CAA47103.1; -;
 DR EMBL; J27090; AAB38270.1; -;
 DR PIR; S26407; S26407.
 DR PIR; JH0800; JH0800.
 DR PDB; 1BE9; 2I-OCT-98.
 DR PDB; 1BE9; 2I-OCT-98.
 DR InterPro; IPR000619; Guanylate_kin.
 DR InterPro; IPR001478; PDZ.
 DR InterPro; IPR001452; SH3.
 DR Pfam; PF00625; Guanylate_kin; 1.
 DR Pfam; PF00595; PDZ; 3.
 DR Pfam; PF00018; SH3; 1.
 DR SMART; SM00072; Gukc; 1.
 DR SMART; SM00228; PDZ; 3.
 DR SMART; SM00326; SH3; 1.
 DR PROSITE; PS00856; GUANYLATE_KINASE_1; 1.
 DR PROSITE; PS00852; GUANYLATE_KINASE_2; 1.
 DR PROSITE; PS50106; PDZ; 3.
 DR PROSITE; PS50002; SH3; 1.
 KW SH3 domain; Repeat; 3D-structure.
 FT DOMAIN 65 151 PDZ 1.
 FT DOMAIN 160 246 PDZ 2.
 FT DOMAIN 313 393 PDZ 3.
 FT DOMAIN 428 498 SH3.
 FT DOMAIN 534 724 SH3.
 FT CONFLICT 61 61 M -> L (IN REF. 2).
 FT CONFLICT 78 78 S -> T (IN REF. 2).
 FT CONFLICT 177 182 GVGNGH -> ALGTSI (IN REF. 2).
 FT CONFLICT 200 200 A -> G (IN REF. 2).
 FT CONFLICT 254 254 S -> T (IN REF. 2).
 FT CONFLICT 539 555 ILPTKDRANDILISEF -> ISLDPKTVPTMIFSSS
 FT (IN REF. 2).
 FT FT GKH -> RDO (IN REF. 3).
 FT CONFLICT 623 625
 FT SEQUENCE 724 AA; 80465 MW; 7922D4BE0F9A085 CRC64;

```

OY      34  GGAAGCTGTLGGLEHGEPLITSKVEEGKADTLSSKLAQAGDEVYHINEYTL-SSSKREAV 92
Db      319  GSTGGLGFINIVOG-EDGEGIFISFILAGSPAD-LSGELRKGDJILSYNGVDLRLNASHQOA 376
OY      93  SLVKGSKYKTLRLAYR 107
Db      377  IALKNAGQTVITIAQ 391

RESULT 15
DLG4_HUMAN STANDARD: PRT; 767 AA.
AC P78352; OS2941; Q9UKK8;
AT 01-NOV-1997 (Rel. 35, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PRESYNAPTIC DENSITY PROTEIN 95 (PSD-95) (DISCS, LARGE HOMOLOG 4).
DLG4 OR PSD95.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_Taxid:9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-Mammary gland;
RX MEDLINE=97432822; PubMed=9286702;
RA Stethakis D.G., Hoover K.B., You Z., Bryant P.J.;
RT "Human postsynaptic density-95 (PSD95): location of the gene (DLG4)
RL and possible function in nonneural as well as in neural tissues.";
RN [2]
RP REVISIONS.
RC TISSUE-Mammary gland;
RL Stethakis D.G., Hoover K.H., You Z., Bryant P.J.;
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=20047407; PubMed=10582582;
RA Stethakis D.G., Udar N., Sandgren O., Andreasson S., Bryant P.J.,
RA Small K., Forsman-Semb K.;
RT "Genomic organization of human DLG4, the gene encoding postsynaptic
RL density 95.";
RL J. Neurochem. 73:2250-2265(1999).
RN [4]
RP SEQUENCE OF 81-401 FROM N.A.
RC TISSUE-Brain;
RA Breman J.H., Bredt D.S., Parkinson J.F., Manzana W.P., Mcclary J.A.;
RL Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.
RL -1- FUNCTION: INTERACTS WITH THE CYTOPLASMIC TAIL OF NMDA RECEPTOR
SUBUNITS. MAY BE INVOLVED IN SYNAPTOGENESIS.
RL -1- SUBCELLULAR LOCATION: CONCENTRATED AT SYNAPTIC JUNCTIONS PRIMARILY
ON THE PRESYNAPTIC SIDE (WAS ORIGINALLY THOUGHT TO BE
PRESYNAPTIC).
CC -1- TISSUE SPECIFICITY: PRESYNAPTIC DENSITY FRACTION OF BRAIN.
CC -1- SIMILARITY: CONTAINS 3 PDZ/DHR DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -1- SIMILARITY: CONTAINS A GUANYLATE KINASE-LIKE DOMAIN.
CC -1- SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb.slb.ch)
CC -----
DR EMBL; 083192; AAC52113.1; -

```

```
DR EMBL; AF156495; AAD56173.1; -.
DR EMBL; U68138; AAB07736.1; -.
DR HSSP; Q12959; IPDR.
DR MIN; 602897; -.
DR InterPro; IPR000619; Guanylate_kin.
DR InterPro; IPR01478; PDZ.
DR InterPro; IPR01452; SH3.
DR Pfam; PF00625; Guanylate_kin; 1.
DR Pfam; PF00595; PDZ; 3.
DR Pfam; PF00018; SH3; 1.
DR SMART; SM00072; GUKC; 1.
DR SMART; SM00228; PDZ; 3.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS00856; GUANYLATE_KINASE_1; 1.
DR PROSITE; PS0052; GUANYLATE_KINASE_2; 1.
DR PROSITE; PS0106; PDZ; 3.
DR PROSITE; PS50002; SH3; 1.
DR SH3 domain; Repeat.
KW DOMAIN 108 194 PDZ 1.
FT DOMAIN 203 289 PDZ 2.
FT DOMAIN 356 436 PDZ 3.
FT DOMAIN 471 541 SH3.
FT DOMAIN 577 767 GUANYLATE_KINASE.
FT CONFLICT 46 46 E -> V (IN REF. 3).
FT CONFLICT 81 83 VIV -> EFR (IN REF. 4).
FT CONFLICT 399 401 GDO -> AGI (IN REF. 4).
SQ SEQUENCE 767 AA; 85429 MW; BE1019159E65B2D8 CRC64;

Query Match 19.2%; Score 104.5; DB 1; Length 767;
Best Local Similarity 34.7%; Pred. No. 0.0068;
Matches 26; Conservative 20; Mismatches 26; Indels 3; Gaps 3;

QY 34 GGAPWGFTIKGGLGEPLIISKVEEGKADTLSSKLOAGDEVVHINEVTL-SSSRKEAV 92
DB 362 GSTGLGFGNIVGG-EDGEGIFISFILAGPAD-LSGELRKGDQILSVNGVDLNRNASHEQAA 419
QY 93 SLVKGSYKTLRLVVR 107
DB 420 IALKNAGQTVIIAQ 434
```

Search completed: January 31, 2002, 21:06:28
Job time: 51 sec

DR N-PSDB; AAS00040.
 PT Novel human polynucleotides isolated from human mammary gland cDNA
 PT library, encodes novel human proteins which are useful in diagnosis,
 PT drug screening, clinical trial monitoring or treating behavioural
 PT disorders
 XX
 XX
 PS Claim 4; Page 25; 29pp; English.
 XX
 CC The sequence is a novel human protein, NHP#2, which shares
 CC structural motifs with human APXL protein. Nucleotide constructs
 CC encoding functional NHPs are used in gene therapy approaches for the
 CC modulation of NHP expression. NHP oligonucleotides can be used as
 CC hybridisation probes for screening libraries and assessing NHP gene
 CC expression patterns. Also, labeled NHP nucleotide probes can be used to
 CC screen a human genomic library. The NHP nucleotide sequences are also
 CC useful in drug screening techniques for treating symptomatic or
 CC phenotypic manifestations of perturbing the normal function of NHP in the
 CC body. Examples of such manifestations may include obesity and high blood
 CC pressure.
 XX
 XX Sequence 108 AA;

Query Match 100.0%; Score 545; DB 22; Length 108;
 Best Local Similarity 100.0%; Pred. No. 1e-59;
 Matches 108; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MMRTEDFHPSATLNSNTATKGRYYLEAFLEGGAPWGFTLKGLEHGEPLIISKVEEG 60
 Db 1 mmrtedfhpsatlnsntatkgryyileafleggapwgftlkgglehgeplliiskveeg 60
 QY 61 GKADTLSSKLAQDEVVHINEVTLSRKEAVSLVKGSKYKTLRLVRS 108
 Db 61 gkadtissklagdevvhinevtlssrkeavslvkgskytclrlvrs 108

RESULT 2
 AAU00031
 ID AAU00031 standard; Protein; 190 AA.
 XX
 AC AAU00031;
 XX
 DT 11-MAY-2001 (first entry)
 XX
 DE Novel human protein, NHP#1.
 XX
 KW Human; novel human protein; NHP#1; gene therapy; drug screening;
 KW obesity; high blood pressure.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Misc_feature 1
 FT /note= "Absent if the true start codon is
 FT nucleotides 4-6 of sequence in AAS00039, the authors
 FT are uncertain which is the true start codon"
 XX
 XX WO200114422-A1.
 XX
 PD 01-MAR-2001.
 XX
 PF 18-AUG-2000; 2000WO-US22815.
 XX
 PR 24-AUG-1999; 99US-0150511.
 XX
 PA (LEXI-) LEXICON GENETICS INC.
 XX
 PI Turner CA, Zambrowicz B, Friedrich G, Nehls M, Sands AT;
 XX
 DR WPI; 2001-218430/22.
 DR N-PSDB; AAS00031.
 XX
 XX Novel human polynucleotides isolated from human mammary gland cDNA
 PT library, encodes novel human proteins which are useful in diagnosis,
 PT drug screening, clinical trial monitoring or treating behavioural
 PT disorders
 XX

PT Novel human polynucleotides isolated from human mammary gland cDNA
 PT library, encodes novel human proteins which are useful in diagnosis,
 PT drug screening, clinical trial monitoring or treating behavioural
 PT disorders
 XX
 XX
 PS Claim 2; Page 24-25; 29pp; English.
 XX
 CC The sequence is a novel human protein, NHP#1, which shares
 CC structural motifs with human APXL protein. Nucleotide constructs
 CC encoding functional NHPs are used in gene therapy approaches for the
 CC modulation of NHP expression. NHP oligonucleotides can be used as
 CC hybridisation probes for screening libraries and assessing NHP gene
 CC expression patterns. Also, labeled NHP nucleotide probes can be used to
 CC screen a human genomic library. The NHP nucleotide sequences are also
 CC useful in drug screening techniques for treating symptomatic or
 CC phenotypic manifestations of perturbing the normal function of NHP in the
 CC body. Examples of such manifestations may include obesity and high blood
 CC pressure.
 XX
 XX Sequence 190 AA;

Query Match 100.0%; Score 545; DB 22; Length 190;
 Best Local Similarity 100.0%; Pred. No. 2.2e-59;
 Matches 108; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MMRTEDFHPSATLNSNTATKGRYYLEAFLEGGAPWGFTLKGLEHGEPLIISKVEEG 60
 Db 1 mmrtedfhpsatlnsntatkgryyileafleggapwgftlkgglehgeplliiskveeg 60
 QY 61 GKADTLSSKLAQDEVVHINEVTLSRKEAVSLVKGSKYKTLRLVRS 108
 Db 61 gkadtissklagdevvhinevtlssrkeavslvkgskytclrlvrs 108

RESULT 3
 AAU00033
 ID AAU00033 standard; Protein; 133 AA.
 XX
 AC AAU00033;
 XX
 DT 11-MAY-2001 (first entry)
 XX
 DE Novel human protein, NHP#3.
 XX
 KW Human; novel human protein; NHP#3; gene therapy; drug screening;
 KW obesity; high blood pressure.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Misc_feature 1
 FT /note= "Absent if the true start codon is
 FT nucleotides 4-6 of sequence in AAS00041, the authors
 FT are uncertain which is the true start codon"
 XX
 XX WO200114422-A1.
 XX
 PD 01-MAR-2001.
 XX
 PF 18-AUG-2000; 2000WO-US22815.
 XX
 PR 24-AUG-1999; 99US-0150511.
 XX
 PA (LEXI-) LEXICON GENETICS INC.
 XX
 PI Turner CA, Zambrowicz B, Friedrich G, Nehls M, Sands AT;
 XX
 DR WPI; 2001-218430/22.
 DR N-PSDB; AAS00041.
 XX
 XX Novel human polynucleotides isolated from human mammary gland cDNA
 PT library, encodes novel human proteins which are useful in diagnosis,
 PT drug screening, clinical trial monitoring or treating behavioural
 PT disorders
 XX

XX FH Key Location/Qualifiers
 FT Domain 91..171
 FT /note= "SH3 domain"
 FT Domain 93..174
 FT /note= "PDZ domain"
 XX WO200142285-A2.
 XX 14-JUN-2001.
 XX 05-DEC-2000; 2000WO-US32990.
 XX 10-DEC-1999; 99US-0172852.
 XX 16-DEC-1999; 99US-0172354.
 XX (INCY-) INCYTE GENOMICS INC.
 XX Yue H, Tang YT, Lal P, Burford N, Azimzai Y, Patterson C;
 XX Baughn MR, Lu DAM, Shah P, Au-Young J;
 XX WPI: 2001-381632/40.
 XX N-PSDB; AAD08063.
 XX New human extracellular matrix and cell adhesion molecules and
 XX polynucleotide sequences encoding them, useful for diagnosis,
 XX prevention, treatment of genetic, autoimmune and cell proliferative
 XX disorders
 XX Claim 1; Page 111-112; 135pp; English.
 XX The present sequence is a human extracellular matrix and cell
 XX adhesion molecule (XNAD). The XNAD is used for screening a compound for
 XX effectiveness as an agonist or antagonist of XNAD. The identified agonist
 XX or antagonist are used for treating a disease or condition associated
 XX with decreased or increased expression of functional XNAD. The
 XX polynucleotides encoding XNAD are useful in somatic or germline gene
 XX therapy to correct a genetic deficiency, to express a conditionally
 XX lethal gene product and to express a protein which affords protection
 XX against intracellular parasites and also for diagnosis of disorders
 XX associated with expression of XNAD. They are also used for generating
 XX hybridisation probes useful in mapping the naturally occurring genomic
 XX sequences and to create knock in humanised animals (pigs) or transgenic
 XX animals (mice or rats) to model human diseases. Oligonucleotide or longer
 XX fragments derived from the polynucleotide sequences may be used as
 XX elements on a microarray. Antibodies which specifically bind XNAD may be
 XX used for the diagnosis of disorders associated with the expression of
 XX XNAD, or in assays to monitor patients being treated with XNAD. Diseases
 XX diagnosed, prevented or treated include genetic disorders such as
 XX adrenoleukodystrophy, Down's syndrome, cystic fibrosis, Gaucher's
 XX disease, myotonic dystrophy, sickle cell anaemia, thalasassaemia,
 XX autoimmune/inflammatory disorders such as acquired immune deficiency
 XX syndrome (AIDS), Addison's disease, allergies, anaemia, asthma,
 XX atherosclerosis, Crohn's disease, diabetes mellitus, atopic dermatitis,
 XX glomerulonephritis, multiple sclerosis, Grave's disease, osteoarthritis,
 XX osteoporosis, psoriasis, rheumatoid arthritis, ulcerative colitis,
 XX bacterial, fungal, parasitic, protozoal and helminthic infections and
 XX cell proliferative disorders such as actinic keratosis, arteriosclerosis
 XX and cancer including breast, bladder, bone marrow, brain and uterus
 XX cancer, leukaemia, adenocarcinoma, lymphoma, melanoma and myeloma.
 XX SQ Sequence 207 AA;

Query Match 22.8%; Score 124.5; DB 22; Length 207;
 Best Local Similarity 36.0%; Pred. No. 3.1e-07;
 Matches 36; Conservative 21; Mismatches 34; Indels 9; Gaps 4;

QY 13 ATNSNTATKNG----RYIYLEAFLEGAPWFTLKGGLHGEPLIISKVEEGKADTLSS 68
 Db 77 atvaatfaseghahprvvpkptdeg---lgfninggkeqnsplivsrpvgvadr-hg 132
 QY 69 KLAGDEVHINEVTLSSSRKE-ANSLVKSGSYKTLRLVVR 107

Db 133 gikrgdqlsvngvsvegeqhekaavellikaagsvklvvr 172
 RESULT 6
 AAW70326
 ID AAW70326 standard; Protein; 406 AA.
 XX AC AAW70326;
 XX 21-DEC-1998 (first entry)
 XX Secreted protein DU123_1.
 XX Secreted protein; DU123_1; human.
 XX Homo sapiens.
 XX Key Location/Qualifiers
 FT Misc-difference 390
 FT /note= "encoded by AAY"
 XX WO9838209-A2;
 XX 03-SEP-1998.
 XX 25-FEB-1998; 98WO-US03697.
 XX 24-FEB-1998; 98US-0028724.
 XX 26-FEB-1997; 97US-0805819.
 XX (GEMY) GENETICS INST INC.
 XX Agostino MJ, Jacobs K, LaVallie ER, McCoy JM, Merberg D;
 XX Racie LA, Spaulding V, Treacy M;
 XX WPI: 1998-481139/41.
 XX N-PSDB; AAV33198.
 XX New isolated polynucleotide(s) and encoded polypeptide(s) -
 XX obtained from human foetal kidney, adult colon, adult brain, foetal
 XX brain and placenta cDNA libraries.
 XX Claim 33; Page 80-81; 103pp; English.
 XX This is the amino acid sequence of novel human secreted protein
 XX DU123_1, as deduced from a full-length cDNA clone (see AAV33198)
 XX obtained from a human foetal brain cDNA library. Database
 XX searching revealed some similarity between DU123_1 and some known
 XX sequences. The invention provides new isolated polynucleotides
 XX (see AAV33198-99), from human foetal kidney, adult colon, adult brain,
 XX foetal brain and placenta cDNA libraries, that code for recombinant
 XX proteins (see AAW70319-27). The clones can be used for recombinant
 XX production of the polypeptides, which may have activities such as
 XX e.g. nutritional activity, cytokine and cell proliferation or
 XX differentiation activity, immunostimulant or immunosuppressive,
 XX haematopoiesis regulating activity, tissue growth activity, activin
 XX or inhibin activity, chemotactic or chemokinetic activity,
 XX haemostatic and thrombolytic activity, receptor/ligand activity,
 XX antiinflammatory activity, cadherin/tumour invasion suppressor
 XX activity, tumour inhibition activity, or other activities.
 XX SQ Sequence 406 AA;

Query Match 22.5%; Score 122.5; DB 19; Length 406;
 Best Local Similarity 36.4%; Pred. No. 1.4e-06;
 Matches 28; Conservative 16; Mismatches 30; Indels 3; Gaps 2;

QY 32 LEGAPAGFTLKGGLHGEPLIISKVEEGKADTLSSKLAGDEVHIN-EVTLSSSRKE 90
 Db 11 mtggapwgrlqggkdqkqplqvakinrnskas--ggglcegdvsvsngnncadltype 68


```

Query Match      22.5%; Score 122.5; DB 20; Length 913;
Best Local Similarity 27.7%; Pred. No. 4.5e-06;
Matches 33; Conservative 23; Mismatches 44; Indels 19; Gaps 4;

QY 6 EDFHKPSATLNSNTATKGRY-----IYLEAFLEGAPWGFTLKGLGHEGEPLI 53
   :|||:  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 479 ddfhr--vtkggstedasqycdkndngdsvlvirtpdgdgkfgnlkgvvdqkmlpv 536
   :|||:  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :

QY 54 ISKYEEGGKADTLSSKLQAGDEVVHINEVTLS--SRKEAVSLVKGSYKT----LRLVVR 107
   :|||:  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 537 vsrinpspadtcipkinegdqvlngdrdisethdqvmvfikasreshreialvir 595
   :|||:  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :

RESULT 9
AAY07095
ID AAY07095 standard; Protein; 403 AA.
XX
XX AAY07095;
XX
XX
XX
XX 02-JUL-1999 (first entry)
XX
XX Colon cancer associated antigen precursor sequence.
XX
XX Cancer associated antigen; diagnosis; research; treatment; human;
KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
KW prostate cancer.
XX
XX Homo sapiens.
XX
XX WO9904265-A2.
PN

```

```

Query Match      22.4%   Score 122; DB 20; Length 403;
Best Local Similarity 41.4%; Pred. No. 1.6e-06;
Matches 29; Conservative 20; Mismatches 17; Indels 4; Gaps 3;

QY    39 GFLLKGLGEGCEPLLISKVEGGADTLSSKLQAGDEVVHINEVTLS-SRKEAVSLVKG 97
       |::|::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::
Db     99 glsvrglgefgcgflshllkggadavg--lvqdeivrngysissctheevinlr- 155
               ::::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::
QY    98 SYKTRLRVLR 107
           :||::|::|
Db     156 tktvtsikvr 165

RESULT 10
AAY07098
ID AAY07098 standard; Protein; 481 AA.
XX AC
XX AC AAY07098;
XX DT
XX XX 02-JUL-1999 (first entry)
XX DE Colon cancer associated antigen precursor sequence.
XX DE Cancer associated antigen; diagnosis; research; treatment; human; breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer; prostate cancer.
XX OS Homo sapiens.
XX
```



```

OS Homo sapiens.
XX OS
XX PN WO9904265-A2.
XX PN
XX PD 28-JAN-1999.
XX PD
XX PF 15-JUL-1998; 98WO-US14679.
XX PF
XX PR 22-JUN-1998; 98US-0102322.
XX PR
XX PR 17-JUL-1997; 97US-0896164.
XX PR
XX PR 10-OCT-1997; 97US-0061599.
XX PR
XX PR 10-OCT-1997; 97US-0061765.
XX PR
XX PR 10-OCT-1997; 97US-0948705.
XX PR
XX PR 11-OCT-1997; 97GB-0021697.
XX PR
XX PA (LUDW-) LUDWIG INST CANCER RES.
XX PA
XX PI Chen Y, Gout I, Gure A, O'Hare M, Obata Y, Old LJ;
XX PI Pfreundschuh M, Sahin U, Scanlan MJ, Stockert E;
XX PI Tureci O;
XX PI
XX DR WPI; 1999-132448/11.
XX DR
XX PT New isolated cancer associated nucleic acids and polypeptides -
XX PT isolated using sera from cancer patients, used to develop products
XX PT for the diagnosis, monitoring or treatment of cancers
XX PT
XX PS Disclosure; Page 656-658; 787pp; English.
XX PS
XX CC The invention relates to a method for diagnosing a disorder characterised
XX CC by expression of a human cancer associated antigen precursor coded for by
XX CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a
XX CC biological sample isolated from a subject with an agent that specifically
XX CC binds to the NAM, an expression product or a fragment of an expression
XX CC product complexed with an HLA molecule; and (b) determining the
XX CC interaction between the agent and the NAM or the expression product as a
XX CC determination of the disorder. The products and methods can be used in
XX CC the diagnosis, monitoring, research, or treatment of conditions
XX CC characterised by the expression of various cancer associated antigens.
XX CC The invention provides nucleic acid sequences and encoded polypeptides
XX CC which are cancer associated antigen precursors expressed in human breast
XX CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
XX CC lung cancer.
XX CC
XX SQ Sequence 652 AA;

Query Match 22.4%; Score 122; DB 20; Length 652;
Best Local Similarity 41.4%; Pred. No. 3.2e-06;
Matches 29; Conservative 20; Mismatches 17; Indels 4; Gaps 3;

QY 39 GFTLKGLGHEGEPLIISKVEEGKADTLSSKLOAGDEVHINEVTLS-SRKEAVSLVKG 97
| : : : | | | | | : : : | | : : : | | : : : | | : : : | | : : : |
Db 99 glsvrgglefvcgflfshllkqgqadsvg--lqvqdeivringysissctheevinlir- 155
QY 98 SYKTLRLVLR 107
: | : : | |
Db 156 tkktvsikvr 165

RESULT 13
AAY07050
ID AAY07050 standard; Protein; 652 AA.
XX AC
XX AC AAY07050;
XX AC
XX DT 02-JUL-1999 (first entry)
XX DT
XX DE Renal cancer associated antigen precursor sequence.
XX DE
XX KW Cancer associated antigen; diagnosis; research; treatment; human;
XX KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
XX KW prostate cancer.

```

XX Synthetic.
 OS
 XX
 PM WO200069896-A2.
 XX
 PD 23-NOV-2000.
 XX
 PF 12-MAY-2000; 2000MO-US13161.
 XX
 PR 14-MAY-1999; 99US-0134114.
 PR 14-MAY-1999; 99US-0134117.
 PR 14-MAY-1999; 99US-0134118.
 PR 21-OCT-1999; 99US-0160860.
 PR 29-OCT-1999; 99US-0162498.
 PR 13-DEC-1999; 99US-0170453.
 PR 14-JAN-2000; 2000US-0176195.
 PR 14-FEB-2000; 2000US-0182296.
 PR 11-APR-2000; 2000US-0196460.
 PR 11-APR-2000; 2000US-0196527.
 XX
 PA (ARBO-) ARBOR VITA CORP.
 XX
 PI Lu PS;
 XX
 DR WPI; 2001-080245/09.
 XX
 PT Modulating a biological function of an endothelial cell or
 PT hematopoietic cell, useful for treating autoimmune diseases and
 PT infectious diseases, by administering an antagonist that inhibits
 PT binding between a PDZ protein and a PL protein -
 XX
 PS Disclosure; Page 28-43; 141pp; English.
 XX
 CC The present invention relates to a new method for modulating a
 CC biological function of an endothelial cell or hematopoietic cell. The
 CC method involves introducing into a cell, an antagonist that inhibits
 CC binding between a PDZ protein and a PL protein. The inhibitor is used
 CC to treat a disease mediated by hematopoietic cells, e.g. autoimmune
 CC disease. It may also be used to prevent transplantation rejection of
 CC a solid organ transplant. The method may also be used in the treatment
 CC of inflammation, allergy, inflammatory bowel diseases, ulcerative
 CC colitis, ileitis, psoriasis, asthma, atopic dermatitis, autoimmune
 CC diseases (e.g. rheumatoid arthritis, multiple sclerosis, osteoarthritis,
 CC insulin-dependent diabetes, Hashimoto thyroiditis, osteoarthritis,
 CC graft rejection, transplant rejection), atherosclerosis, cancers,
 CC infectious diseases, ischemia, vasculitis and Crohn's disease.
 CC
 CC Sequence 86 AA;
 XX
 SQ
 Query Match 21.6%; Score 117.5; DB 22; Length 86;
 Best Local Similarity 35.1%; Pred. No. 6,6e-07;
 Matches 27; Conservative 18; Mismatches 29; Indels 3; Gaps 2;
 OY 32 LBSGAPMGFTIKGLHGEPLITSKVEBGKADTLSSKLOAGEVHINEVTLSS-SRKE 90
 DB 8 lvgpawgfrlrggkdfmptlssikdgka--agavrriqdvavlsdginaqgmthle 65
 OY 91 AVSLVGSYKTLRLVVR 107
 DB 66 aguklkycctslmrtiq 82
 XX
 AC AAB57631;
 XX
 DT 12-MAR-2001 (first entry)
 XX
 DE Human LIM protein PDZ domain.
 XX

KW Endothelial cell; hematopoietic cell; PDZ domain protein;
 KW PL domain protein; leukocyte activation; synapse formation;
 KW transmembrane neurotransmitter receptor; autoimmune disease;
 KW transplantation rejection; inflammation; allergy;
 KW inflammatory bowel disease; ulcerative colitis; ileitis; psoriasis;
 KW asthma; atopic dermatitis; atherosclerosis; cancer; infectious disease;
 KW ischemia; vasculitis; Crohn's disease.
 XX
 OS Homo sapiens.
 XX
 PN WO200069897-A2.
 XX
 PD 23-NOV-2000.
 XX
 PF 12-MAY-2000; 2000MO-US13166.
 XX
 PR 14-MAY-1999; 99US-0134114.
 PR 14-MAY-1999; 99US-0134117.
 PR 14-MAY-1999; 99US-0134118.
 PR 21-OCT-1999; 99US-0160860.
 PR 29-OCT-1999; 99US-0162498.
 PR 13-DEC-1999; 99US-0170453.
 PR 14-JAN-2000; 2000US-0176195.
 PR 14-FEB-2000; 2000US-0182296.
 PR 11-APR-2000; 2000US-0196460.
 PR 11-APR-2000; 2000US-0196527.
 XX
 PA (ARBO-) ARBOR VITA CORP.
 XX
 PI Lu PS;
 XX
 DR WPI; 2001-025003/03.
 XX
 PT New inhibitors of binding of a PDZ protein and PL protein for
 PT inhibiting T cell-mediated response by hematopoietic cells, or for
 PT treating diseases characterized by inflammatory and humoral immune
 PT responses, e.g. inflammation, cancer -
 XX
 PS Disclosure; Page 32; 139pp; English.
 XX
 CC The present invention relates to a method for modulating a biological
 CC function of an endothelial cell or hematopoietic cell, comprises
 CC introducing into a cell an antagonist that inhibits binding between a
 CC PDZ domain protein and a PL domain protein to result in inhibition of
 CC leukocyte activation. The present sequence is a PDZ domain. PDZ domains
 CC of proteins are named after three prototypical proteins: PSD95,
 CC Drosophila large disc protein and Zonula Occludin 1 protein. PDZ domain
 CC proteins are involved in synapse formation by organising transmembrane
 CC neurotransmitter receptors through intracellular interactions. The
 CC inhibitors identified by the present invention can be used to treat a
 CC disease mediated by hematopoietic cells, e.g. autoimmune disease,
 CC inflammation, allergy (e.g. drug allergies), inflammatory bowel diseases,
 CC ulcerative colitis, ileitis, psoriasis, respiratory allergic diseases
 CC (e.g. asthma), atopic dermatitis, autoimmune diseases (e.g. rheumatoid
 CC arthritis, multiple sclerosis, insulin-dependent diabetes, Hashimoto
 CC thyroiditis, osteoarthritis), atherosclerosis, cancers, infectious
 CC diseases (e.g. viral infection), ischemia, vasculitis and Crohn's
 CC disease. The inhibitors can also be used to prevent transplantation
 CC rejection of a solid organ transplant.
 CC
 CC Sequence 86 AA;
 XX
 SQ
 Query Match 21.6%; Score 117.5; DB 22; Length 86;
 Best Local Similarity 35.1%; Pred. No. 6,6e-07;
 Matches 27; Conservative 18; Mismatches 29; Indels 3; Gaps 2;
 OY 32 LBSGAPMGFTIKGLHGEPLITSKVEBGKADTLSSKLOAGEVHINEVTLSS-SRKE 90
 DB 8 lvgpawgfrlrggkdfmptlssikdgka--agavrriqdvavlsdginaqgmthle 65
 OY 91 AVSLVGSYKTLRLVVR 107
 DB 66 aguklkycctslmrtiq 82
 XX

Db 66 agnkikgetgslnmtlq 82

Search completed: January 31, 2002, 20:56:12
Job time: 80 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 20:57:10 ; Search time 29.66 Seconds
(without alignments)
81.941 Million cell updates/sec

Title: US-09-641-831-4

Perfect score: 545
Sequence: 1 MMRTEDFHKPSATLNSNTA.....KEAVSLVKGSKYTLRLVRS 108

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

1: Issued_Patents_AA.*
2: /cgn2_6/ptodata/2/1aa/5A.COMB.pep.*
3: /cgn2_6/ptodata/2/1aa/5B.COMB.pep.*
4: /cgn2_6/ptodata/2/1aa/6A.COMB.pep.*
5: /cgn2_6/ptodata/2/1aa/6B.COMB.pep.*
6: /cgn2_6/ptodata/2/1aa/6C.COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	117	21.5	75	US-09-100-804-21	Sequence 21, Appl
2	113.5	20.8	327	US-08-739-485-10	Sequence 10, Appl
3	112.5	20.6	233	US-09-151-611-1	Sequence 1, Appl
4	112.5	20.6	233	US-09-370-102-1	Sequence 1, Appl
5	112.5	20.6	329	US-08-739-485-3	Sequence 3, Appl
6	112.5	20.6	330	US-08-739-485-11	Sequence 11, Appl
7	109.5	20.1	316	US-08-739-485-5	Sequence 5, Appl
8	109	20.0	297	US-09-151-611-3	Sequence 3, Appl
9	109	20.0	297	US-09-370-102-3	Sequence 3, Appl
10	106.5	19.5	223	US-09-124-238A-34	Sequence 34, Appl
11	106.5	19.5	457	US-09-124-238A-1	Sequence 1, Appl
12	106.5	19.5	457	US-09-124-238A-10	Sequence 10, Appl
13	104.5	19.2	454	US-08-166-316-2	Sequence 2, Appl
14	103.5	19.0	73	US-09-100-804-28	Sequence 28, Appl
15	95	17.4	182	US-09-045-632-29	Sequence 29, Appl
16	95	17.4	283	US-09-045-632-27	Sequence 27, Appl
17	95	17.4	498	US-09-045-632-30	Sequence 30, Appl
18	95	17.4	502	US-09-045-632-24	Sequence 24, Appl
19	95	17.4	505	US-08-123-161A-14	Sequence 14, Appl
20	95	17.4	505	US-08-483-278-14	Sequence 14, Appl
21	95	17.4	541	US-09-045-632-36	Sequence 36, Appl
22	95	17.4	599	US-09-045-632-28	Sequence 28, Appl
23	95	17.4	602	US-09-045-632-20	Sequence 20, Appl
24	95	17.4	642	US-09-045-632-35	Sequence 35, Appl
25	95	17.4	702	US-09-045-632-15	Sequence 15, Appl
26	95	17.4	818	US-09-045-632-25	Sequence 25, Appl
27	95	17.4	861	US-09-045-632-34	Sequence 34, Appl

28	95	17.4	918	3	US-09-045-632-21	Sequence 21, Appl
29	95	17.4	961	3	US-09-045-632-33	Sequence 33, Appl
30	95	17.4	1018	3	US-09-045-632-16	Sequence 16, Appl
31	95	17.4	1061	3	US-09-045-632-32	Sequence 32, Appl
32	95	17.4	1112	3	US-09-045-632-2	Sequence 2, Appl
33	95	17.4	1112	3	US-09-045-632-3	Sequence 3, Appl
34	93.5	17.2	80	3	US-08-545-860D-54	Sequence 54, Appl
35	93.5	17.2	80	5	PCT-US94-04496-54	Sequence 54, Appl
36	92.5	17.0	73	3	US-09-100-804-25	Sequence 25, Appl
37	92.5	17.0	74	3	US-09-100-804-34	Sequence 34, Appl
38	88	16.1	82	3	US-09-045-632-9	Sequence 9, Appl
39	88	16.1	398	3	US-09-045-632-31	Sequence 31, Appl
40	88	16.1	441	3	US-09-045-632-37	Sequence 37, Appl
41	87.5	16.1	75	3	US-09-100-804-22	Sequence 22, Appl
42	87.5	16.1	79	3	US-09-100-804-27	Sequence 27, Appl
43	87.5	16.1	86	3	US-08-545-860D-53	Sequence 53, Appl
44	87.5	16.1	86	5	PCT-US94-04496-53	Sequence 53, Appl
45	84	15.4	2037	4	US-09-306-998-3	Sequence 3, Appl

ALIGNMENTS

RESULT 1
US-09-100-804-21
Sequence 21, Application US/09100804
Patent No. 6066472
GENERAL INFORMATION:
APPLICANT: GONZ, LEONEL JORGE
APPLICANT: SARAS, JAN
APPLICANT: CLAESON-WELSH, LENA
APPLICANT: HELDIN, CARL-HENRIK
TITLE OF INVENTION: PRIMARY STRUCTURE AND FUNCTIONAL
TITLE OF INVENTION: EXPRESSION OF NOCOTINOTIDE SEQUENCES FOR NOVEL PROTEIN
NUMBER OF SEQUENCES: 34
CORRESPONDENCE ADDRESS:
ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.
STREET: 600 ATLANTIC AVENUE
CITY: BOSTON
STATE: MASSACHUSETTS
COUNTRY: USA
ZIP: 02210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/100,804
FILING DATE:
CLASSIFICATION:
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/596,291
FILING DATE: 09-AUG-1996
APPLICATION NUMBER: US 08/115,573
FILING DATE: 01-SEP-1993
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/09943
FILING DATE: 01-SEP-1994
ATTORNEY/AGENT INFORMATION:
NAME: GATES, EDWARD R.
REGISTRATION NUMBER: 31,616
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-720-3500
TELEFAX: 617-720-2441
TELEX:
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 75 amino acids
TYPE: amino acid
STRANDEDNESS: single

```

; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-09-100-804-21

```

Query Match	21.5%;	Score 117;	DB 3;	Length 75;
Best Local Similarity	38.8%;	Pred. No. 1.3e-07;		
Matches 26;	Conservative 12;	Mismatches 27;	Indels	

QY 33 EGGAPWGFTLKGGLIEHGEPLIISKVEEGKADTLSSKLQAGDEVVHHINEVTLSS-SRKEA 91
| | | | | : : : : : ||| || |||::||| : : :
DB ? FDCGR-CENIGGVGVDOKNPTVYSRIINSPSPADTCIPKINEGDIVLINGRDISEHTHDV 60

Qy 92 VSLVKS 98
| : | |
Db 61 VMFIKAS 67

RESULT 2
US-08-739-485-10
; Sequence 10, Application US/08739485
; Patent No. 5863898
; GENERAL INFORMATION:
; APPLICANT: Goli, Surya K.
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Bandman, Olga
; TITLE OF INVENTION: NOVEL HUMAN LIM PROTEINS
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

INFORMATION FOR SEQ ID NO: 10:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 327 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 IMMEDIATE SOURCE:
 LIBRARY: Genbank
 CLONE: 1020151
 US-08-739-485-10

Query Match 20.8%; Score 113.5; DB 2; Length 327;
Best Local Similarity 39.5%; Pred. No. 3.1e-06;
Matches 30: Conservative 11; Mismatches 32; Indels 3

[illegible]

```

RESULT      3
US-09-151-611-1
: Sequence 1, Application US/09151611
: Patent No. 5958731
: GENERAL INFORMATION:
: APPLICANT: Yue, Henry
: APPLICANT: Au-Young, Janice
: APPLICANT: Patterson, Chandra
: TITLE OF INVENTION: CELL JUNCTION
: FILE REFERENCE: PF-0599 US
: CURRENT APPLICATION NUMBER: US/09/
: CURRENT FILING DATE: 1998-09-11
: NUMBER OF SEQ ID NOS: 3
: SOFTWARE: PERL Program
: SEQ ID NO 1
: LENGTH: 233
: TYPE: PRT
: ORGANISM: Homo sapiens
: FEATURE: -
: OTHER INFORMATION: 1974337
US-09-151-611-1

```

Query Match 20.6%; Score 112.5; DB 2; Length 233;
Best Local Similarity 33.0%; Pred. No. 2.5e-06;
Matches 33; Conservative 22; Mismatches 36; Indels 9

13 ATLSNTRATKG---RYIYLEAFLEGAGPQFTLLKGLEHCEPLIIISKVEGGKADTLLS 69
 92 ATVAFAASEAGSHRPVVPELPTDEG---LGFNMVGGKQNSPIYISRIIPGVAER-HG 147
 69 KIQAGDEVVHNEVTLSSSRKE-AVSLVKGSYKTLRLVVR 107
 148 GIKRCDIIVLVNGSVSVEGEHEKAVELLKAAKDSVKLVVR 187

```

RESULT      4
; US-09-370-102-1
; Sequence 1, Application US/09370102
; Patent No. 6265547
; GENERAL INFORMATION:
; APPLICANT: Yue, Henry
; APPLICANT: Au-Young, Janice
; APPLICANT: Patterson, Chandra
; TITLE OF INVENTION: CELL JUNCTION PDZ PROTEIN
; FILE REFERENCE: PF-0599 US
; CURRENT APPLICATION NUMBER: US/09/370,102
; CURRENT FILING DATE: 1999-08-06
; EARLIER APPLICATION NUMBER: 09/151,611
; EARLIER FILING DATE: 1998-09-11
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PERL Program
; SEQ ID NO 1
; LENGTH: 233
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE: -
; OTHER INFORMATION: 1974337
; US-09-370-102-1

```

Query Match	20.6%;	Score 112.5;	DB 4;	Length 233;
Best Local Similarity	33.0%;	Pred. No. 2.5e-06;		
Matches 33: Conservative	22;	Mismatches 36;	Indels 9;	Gaps 4;

QY 13 ATINSNTATKCG---RYIYLEAFLEGGAPMGFTLKGLEHGEPLITSKVEEGKADTLSS 68
DB 92 ATVAAPFAASEGSHSPRVVLLPTDEG---LGFVVMGKQNSPIYISRLIPGVAAER-HQ 147
QY 69 KQAGDEVHINEVTLSSRKE-AVSLVKGSKYKTLRLV 107
DB 148 GLKRGDOLLVNGSVGEHHEKAVELLKAADSVKLV 187

RESULT 5

US-08-739-485-3
; Sequence 3, Application US/08739485
; Patent No. 5863898
; GENERAL INFORMATION:
; APPLICANT: Goll, Surya K.
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Bandman, Olga
; TITLE OF INVENTION: NOVEL HUMAN LIM PROTEINS
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: US
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/739,485
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0142 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 329 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: Consensus
; CLONE: Consensus
; US-08-739-485-3

Query Match 20.6%; Score 112.5; DB 2; Length 329;
Best Local Similarity 39.5%; Pred. No. 4.2e-06;
Matches 30; Conservative 11; Mismatches 32; Indels 3; Gaps 2;

QY 32 LEGAPMGFTLKGLEHGEPLITSKVEEGKADTLSSKLAGDEVHIN-EVTLSSRKE 90
DB 8 LQGGPMGFRLVGGKDEQPLAISRVIPGSKA--ALANLCIGDVIYTAIDGENTSNMTHLE 65
QY 91 AVSLVKGSKYKTLRLV 106
DB 66 AONRIKCGDHLTLV 81

RESULT 6

US-08-739-485-11
; Sequence 11, Application US/08739485
; Patent No. 5863898
; GENERAL INFORMATION:
; APPLICANT: Goll, Surya K.
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Bandman, Olga
; TITLE OF INVENTION: NOVEL HUMAN LIM PROTEINS
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: US
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/739,485
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0142 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 330 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; ORIGINAL SOURCE:
; HAPLOTYPE: GenBank
; IMMEDIATE SOURCE:
; LIBRARY: 887580
; US-08-739-485-11

Query Match 20.6%; Score 112.5; DB 2; Length 330;
Best Local Similarity 40.8%; Pred. No. 4.2e-06;
Matches 31; Conservative 9; Mismatches 33; Indels 3; Gaps 2;

QY 32 LEGAPMGFTLKGLEHGEPLITSKVEEGKADTLSSKLAGDEVHIN-EVTLSSRKE 90
DB 7 LRGPSPMGFRLVGGKDEQPLAISRVIPGSKA--ALANLCIGDVIYTAIDGENTSNMTHLE 64
QY 91 AVSLVKGSKYKTLRLV 106
DB 65 AONRIKCGDHLTLV 80

RESULT 7
US-08-739-485-5
; Sequence 5, Application US/08739485
; Patent No. 5863898
; GENERAL INFORMATION:
; APPLICANT: Goll, Surya K.
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Bandman, Olga
; TITLE OF INVENTION: NOVEL HUMAN LIM PROTEINS
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:

ADDRESSEE: Incyte Pharmaceuticals, Inc.

STREET: 3174 Porter Drive

CITY: Palo Alto

STATE: CA

COUNTRY: US

ZIP: 94304

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/739,485

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Billings, Lucy J

REGISTRATION NUMBER: 36,749

REFERENCE/DOCKET NUMBER: PF-0142 US

TELECOMMUNICATION INFORMATION:

TELEPHONE: 415-855-0555

TELEFAX: 415-845-4166

TELEX:

INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:

LENGTH: 316 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

IMMEDIATE SOURCE:

LIBRARY: Consensus

CLONE: Consensus

US-08-739-485-5

Query Match

Best Local Similarity 20.1%; Score 109.5; DB 2; Length 316;

Matches 26; Conservative 17; Mismatches 30; Indels 3; Gaps 2;

QY 32 LEGGAPWGTLLKGLHGEPLIISKVEGGKADTLSSKLQAGDEVVHINEV-TLSSSRKE 90

DB 7 LCPAPWGTLLKGLHGEPLIISKVEGGKADTLSSKLQAGDEVVHINEV-TLSSSRKE 90

QY 91 AVSLVKGSKYKTLRLV 106

DB 65 AQDRIKAAHQCLKI 80

RESULT 8

US-09-151-611-3

Sequence 3, Application US/09151611

Patent No. 5958731

GENERAL INFORMATION:

APPLICANT: Yue, Henry

APPLICANT: Au-Young, Janice

APPLICANT: Patterson, Chandra

TITLE OF INVENTION: CELL JUNCTION PDZ PROTEIN

FILE REFERENCE: PF-0599 US

CURRENT APPLICATION NUMBER: US/09/151,611

CURRENT FILING DATE: 1998-09-11

NUMBER OF SEQ ID NOS: 3

SOFTWARE: PERL Program

SEQ ID NO 3

LENGTH: 297

TYPE: PRT

ORGANISM: Caenorhabditis elegans

FEATURE: -

OTHER INFORMATION: g1685067

US-09-151-611-3

Query Match

Best Local Similarity 20.0%; Score 109; DB 2; Length 297;

Matches 30; Conservative 23; Mismatches 38; Indels 8; Gaps 3;

QY 13 ATLNSTATKG---RYIYLEAFLEGGAPWGTLLKGLHGEPLIISKVEGGKADTLSS 68

DB 184 ATVAAPAAAEHGHAPRIVELPKTDG---LGFNVMGKQNSPIYISRIIPGGVADR-HG 239

QY 69 KIQAGDEVVHINEVTLSRKEAVSLVKGSKYKTLRLV 107

DB 240 GLKRGDQLIAVNGVNEAECHERKAVDLLKSAVGSVKLVIR 278

RESULT 9

US-09-370-102-3

Sequence 3, Application US/09370102

Patent No. 6265547

GENERAL INFORMATION:

APPLICANT: Yue, Henry

APPLICANT: Au-Young, Janice

APPLICANT: Patterson, Chandra

TITLE OF INVENTION: CELL JUNCTION PDZ PROTEIN

FILE REFERENCE: PF-0599 US

CURRENT APPLICATION NUMBER: US/09/370,102

CURRENT FILING DATE: 1998-08-06

EARLIER APPLICATION NUMBER: 09/151,611

EARLIER FILING DATE: 1998-09-11

NUMBER OF SEQ ID NOS: 3

SOFTWARE: PERL Program

SEQ ID NO 3

LENGTH: 297

TYPE: PRT

ORGANISM: Caenorhabditis elegans

FEATURE: -

OTHER INFORMATION: g1685067

US-09-370-102-3

Query Match

Best Local Similarity 20.0%; Score 109; DB 4; Length 297;

Matches 30; Conservative 23; Mismatches 38; Indels 8; Gaps 3;

QY 13 ATLNSTATKG---RYIYLEAFLEGGAPWGTLLKGLHGEPLIISKVEGGKADTLSS 68

DB 184 ATVAAPAAAEHGHAPRIVELPKTDG---LGFNVMGKQNSPIYISRIIPGGVADR-HG 239

QY 69 KIQAGDEVVHINEVTLSRKEAVSLVKGSKYKTLRLV 107

DB 240 GLKRGDQLIAVNGVNEAECHERKAVDLLKSAVGSVKLVIR 278

RESULT 10

US-09-124-238A-34

Sequence 34, Application US/09124238A

Patent No. 6300127

GENERAL INFORMATION:

APPLICANT: Hair, Gregory A.

APPLICANT: Boden, Scott D.

TITLE OF INVENTION: No. 6300127el Bone Mineralization Proteins, DNA, Vectors,

FILE REFERENCE: Expression Systems

FILE REFERENCE: 06148.0115

CURRENT APPLICATION NUMBER: US/09/124,238A

CURRENT FILING DATE: 1998-07-29

PRIOR APPLICATION NUMBER: 60/054,219

PRIOR FILING DATE: 1997-07-30

PRIOR APPLICATION NUMBER: 60/080,407

PRIOR FILING DATE: 1998-04-02

NUMBER OF SEQ ID NOS: 36

SOFTWARE: MS Word

SEQ ID NO 34

LENGTH: 223

TYPE: PRT

A:Title: Isolation of a cDNA clone encoding a human protein-tyrosine phosphatase with homology to SH-PTPase
 A:Reference number: A41109; MUID:91296738
 A:Accession: A41109
 A:Molecule type: mRNA
 A:Residues: 1-913 <YAN>
 A:Cross-references: GB:M64572; NID:g179912; PIDN:AAA35647.1; PID:g179913
 R:Ikuota, S.; Itoh, F.; Hinoda, Y.; Toyota, M.; Makiguchi, Y.; Imai, K.; Yachi, A.
 J. Gastroenterol. 29, 727-732, 1994
 A:Title: Expression of cytoskeletal-associated protein tyrosine phosphatase PTPH1 mRNA

RESULT 6
JEO209
brain-specific angiogenesis inhibitor-associated protein 1 - human
N alternate names: BA11-associated protein 1; BAP1 (misnomer)


```
RESULT 10
I38757
homolog of Drosophila discs large protein, isoform 1 - human
C:Species: Homo sapiens (man)
C>Date: 01-Mar-1996 #sequence_revision 01-Mar-1996 #text_change 21-Jan-2000
C:Accession: I38757
R:Lue, R.A.; Marfatia, S.M.; Branton, D.; Chishti, A.H.
Proc. Natl. Acad. Sci. U.S.A. 91, 9818-9822, 1994
A:Title: Cloning and characterization of hdlg: the human homologue of the Drosophila disc
A:Reference number: I38756; MUID:95024052
A:Accession: I38757
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-904 <RES>
A:Cross-references: EMBL:U13897; NID:q558437; PIDN:AAA50599.1; PID:q558438
C:Genetics:
A:Gene: GDB:DLG1
A:Cross-references: GDB:393278; OMIM:601014
A:Map position: 3q29-3q29
C:Superfamily: discs-large tumor suppressor; GLGF domain homology; guanylate kinase hom
C:Keywords: alternative splicing; duplication
F:229-307/Domain: GLGF domain homology <GLG1>
F:324-402/Domain: GLGF domain homology <GLG2>
F:588-646/Domain: SH3 homology <SH3>
F:715-892/Domain: guanylate kinase homology <GKI>

Query Match 20.4%; Score 111; DB 2; Length 904;
Best Local Similarity 30.3%; Pred. No. 0.0021;
Matches 36; Conservative 23; Mismatches 40; Indels 20; Gaps 6;

QY 7 DPH-KPSATLNSNTATKGRYYL-EAFL-----EGGAPWGFTLKGGLHG 49
DB 428 DNHVSPSSFLGQTSPASPARYSPVSKAVLGDDDEITREPRKVVHLRSGTGLGFNIYVG-EDG 486
QY 50 EPLIISKVEEGKADTLSSKLOAGDEVVHINEVTLS-SSSRKEAVSLVKGSKYKTLRLVVR 107
DB 487 EGIFISFILAGGPAD-LSGELRKGDRIISVNSVDLRAASHEQAAAALKNAGQAVTIQAQ 544

RESULT 11
I38756
homolog of Drosophila discs large protein, isoform 2 - human
C:Species: Homo sapiens (man)
C>Date: 01-Mar-1996 #sequence_revision 01-Mar-1996 #text_change 21-Jan-2000
C:Accession: I38756
R:Lue, R.A.; Marfatia, S.M.; Branton, D.; Chishti, A.H.
Proc. Natl. Acad. Sci. U.S.A. 91, 9818-9822, 1994
A:Title: Cloning and characterization of hdlg: the human homologue of the Drosophila dis
A:Reference number: I38756; MUID:95024052
A:Accession: I38756
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-926 <RES>
A:Cross-references: EMBL:U13896; NID:q558435; PIDN:AAA50598.1; PID:q558436
C:Genetics:
A:Gene: GDB:DLG1
A:Cross-references: GDB:393278; OMIM:601014
A:Map position: 3q29-3q29
C:Superfamily: discs-large tumor suppressor; GLGF domain homology; guanylate kinase hom
C:Keywords: alternative splicing; duplication
F:229-307/Domain: GLGF domain homology <GLG1>
F:324-402/Domain: GLGF domain homology <GLG2>
F:588-646/Domain: SH3 homology <SH3>
F:737-914/Domain: guanylate kinase homology <GKI>

Query Match 20.4%; Score 111; DB 2; Length 926;
Best Local Similarity 30.3%; Pred. No. 0.0021;
Matches 36; Conservative 23; Mismatches 40; Indels 20; Gaps 6;

QY 7 DPH-KPSATLNSNTATKGRYYL-EAFL-----EGGAPWGFTLKGGLHG 49
```

```
Db 428 DNHVSPSSFLGQTSPASPARYSPVSKAVLGDDDEITREPRKVVHLRSGTGLGFNIYVG-EDG 486
QY 50 EPLIISKVEEGKADTLSSKLOAGDEVVHINEVTLS-SSSRKEAVSLVKGSKYKTLRLVVR 107
DB 487 EGIFISFILAGGPAD-LSGELRKGDRIISVNSVDLRAASHEQAAAALKNAGQAVTIQAQ 544

RESULT 12
T27179
hypothetical protein Y54G11A.10 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T27179
R:Wallis, J.
submitted to the EMBL Data Library, December 1998
A:Reference number: Z20322
A:Accession: T27179
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-317 <WIL>
A:Cross-references: EMBL:AL034488; NID:el359895; PIDN:CAA22459.1; CBSP:Y54G11A.10
A:Experimental source: clone Y54G11A
C:Genetics:
A:Gene: CBSP:Y54G11A.10
A:Introns: 55/3; 108/3; 175/3; 228/2; 253/3

Query Match 20.3%; Score 110.5; DB 2; Length 317;
Best Local Similarity 31.0%; Pred. No. 0.00066;
Matches 31; Conservative 24; Mismatches 36; Indels 9; Gaps 4;

QY 13 ATLNSNTATK-----RYIYLEAFLEGAPWGFTLKGGLHGEPLIISKVEEGKADTLSS 68
DB 184 ATVAFAAAEGHAHPRIPELPTDQ---LGFNVMGKQKQNSPIYISRIIPGGVADR-HG 239
QY 69 KIQAGDEVVHINEVTLS-SRKEAVSLVKGSKYKTLRLVVR 107
DB 240 GLKRGDQLIANGVNVVEAECEKAKVDLLAKSANGSVKLVR 279

RESULT 13
A55050
enigma - human
C:Species: Homo sapiens (man)
C>Date: 18-Nov-1994 #sequence_revision 18-Nov-1994 #text_change 21-Jul-2000
C:Accession: A55050
R:Wu, R.Y.; Gill, G.N.
J. Biol. Chem. 269, 25085-25090, 1994
A:Title: LIM domain recognition of a tyrosine-containing tight turn.
A:Reference number: A55050; MUID:95014287
A:Accession: A55050
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-455 <WUA>
A:Cross-references: GB:I35240; NID:q561636; PIDN:AAC37565.1; PID:q561637
C:Superfamily: LIM metal-binding repeat homology; GLGF domain homology
F:9-81/Domain: GLGF domain homology <GLG>
F:280-330/Domain: LIM metal-binding repeat homology <LIM1>
F:339-389/Domain: LIM metal-binding repeat homology <LIM2>
F:398-450/Domain: LIM metal-binding repeat homology <LIM3>

Query Match 19.5%; Score 106.5; DB 2; Length 455;
Best Local Similarity 44.4%; Pred. No. 0.0027;
Matches 28; Conservative 7; Mismatches 19; Indels 9; Gaps 2;

QY 32 LEGAPWGFTLKGGLHGEPLIISKVEEGKADTLSSKLOAGDEVVHINEVTLS-SRKEA 91
DB 8 LEGPAPWGFRLOGKDFNVPISRLTPGKA-----AQAG---VAVGDMVLSDGENA 58
QY 92 VSL 94
```

Db 59 GSL 61

Job time: 65 sec

RESULT 14

T47134

hypothetical protein DKFZp76112312.1 - human

C:Species: Homo sapiens (man)

C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000

C:Accession: T47134

R:Ansoyge, W.; Wilkner, U.; Mewes, H.W.; Well, B.; Wiemann, S.

submitted to the Protein Sequence Database, March 2000

A:Reference number: Z24375

A:Accession: T47134

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-517 <AAA>

A:Cross-references: EMBL:AL161971

A:Experimental source: adult amygdala; clone DKFZp76112312

C:Genetics:

A>Note: DKFZp76112312.1

Query Match 19.4%; Score 105.5; DB 2; Length 517;

Best Local Similarity 26.5%; Pred. No. 0.004; Mismatches 37; Indels 15; Gaps 3;

Matches 27; Conservative

QY 6 EDFHPSATLNSNTATKGRYILEAFLEGGAPWFTLKGLEHGEPLIISKVEGGKADT 65

DB 49 EPFYSGERTVTRRQTVGFE-----GLSIKGAEHNPVYVSKISKEGRAE 94

QY 66 ISSKIQAGDEVVHINEVTLSSSR-KEAVSLVKGSTKTLRLV 106

DB 95 LSGLEFGDAILQINGINVRKCRHEEVQVLRNAGEVTLTV 136

RESULT 15

A45436

synapse-associated protein SAP90 - rat

C:Species: Rattus norvegicus (Norway rat)

C:Date: 22-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 21-Jan-2000

C:Accession: A45436

R:Kistner, U.; Menzel, B.M.; Voh, R.W.; Cases-Ianghoff, C.; Garner, A.M.; Appeltauer, U.

J. Biol. Chem. 268, 4580-4583, 1993

A:Title: SAP90, a rat presynaptic protein related to the product of the Drosophila tumor

A:Reference number: A45436; MUID:93186749

A:Accession: A45436

A:Status: preliminary; not compared with conceptual translation

A:Molecule type: nucleic acid

A:Residues: 1-720 <KIS>

A:Experimental source: brain

A>Note: sequence extracted from NCBI backbone (NCBIP:126555)

C:Superfamily: discs-large tumor suppressor; GIGF domain homology; guanylate kinase hom

F:70-148/Domain: GIGF domain homology <GIG1>

F:165-243/Domain: GIGF domain homology <GIG2>

F:435-493/Domain: SH3 homology <SH3>

F:531-708/Domain: guanylate kinase homology <GKI>

Query Match 19.2%; Score 104.5; DB 2; Length 720;

Best Local Similarity 34.7%; Pred. No. 0.0076;

Matches 26; Conservative 20; Mismatches 26; Indels 3; Gaps 3;

QY 34 GGAPWFTLKGLEHGEPLIISKVEGGKADTSSKIQAGDEVVHINEVTL-SSSRKEAV 92

DB 319 GSTGLGFNIYVG-EDRGIRFTLAGGPAD-LSGELRKGDQILSYNGVDLRNASHQAA 376

QY 93 SLVKGSTKTLRLVVR 107

DB 377 IALKNAGQTVTLIAQ 391

Search completed: January 31, 2002, 20:59:17

us-09-641-831-4.rpr

Fri Feb 1 08:56:32 2002

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 21:05:34 ; Search time 72.88 Seconds
(without alignments)
216.759 Million cell updates/sec

Title: US-09-641-831-4

Perfect score: 545
Sequence: 1 MRRTEDEHFRPSATLNSNTA.....KEAVSLYKSGYKTLRLVYRS 108

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08
Maximum Match 100%

Listing first 45 summaries

Database :
1: SP:archaea:*
2: SP:bacteria:*
3: SP:fungi:*
4: SP:human:*
5: SP_invertebrate:*
6: SP_mammal:*
7: SP_mhc:*
8: SP_Organelle:*
9: SP_phage:*
10: SP_plant:*
11: SP_protent:*
12: SP_virus:*
13: SP_vertibrate:*
14: SP_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	409	75.0	1986	11	Q9GXN0 mus musculus
2	262.5	48.2	1502	4	Q9ULH8 Q9ULH8 homo sapien
3	134.5	24.7	1454	4	Q9H4G2 Q9H4G2 homo sapien
4	128	23.5	39	4	Q9H4T0 Q9H4T0 homo sapien
5	125.5	23.0	298	13	Q9PU46 Q9PU46 gallus gall
6	124.5	22.8	207	4	Q9HAP6 Q9HAP6 homo sapien
7	123.5	22.7	207	11	Q9Z2S2 Q9Z2S2 rattus norv
8	123.5	22.7	207	11	Q889S1 Q889S1 mus musculus
9	123.5	22.7	234	11	Q9D2Q1 Q9D2Q1 mus musculus
10	123.5	22.7	315	13	Q9PU47 Q9PU47 gallus gall
11	123	22.6	171	5	017458 017458 schistosoma
12	122	22.4	403	4	Q9UPC3 Q9UPC3 homo sapien
13	122	22.4	548	11	Q9ES65 Q9ES65 mus musculus
14	122	22.4	552	4	Q9UM04 Q9UM04 homo sapien
15	122	22.4	652	4	Q9Y6N9 Q9Y6N9 mus musculus
16	122	22.4	910	11	Q9ES64 Q9ES64 mus musculus
17	121	22.2	493	5	Q9W443 Q9W443 drosophila
18	120	22.0	521	4	Q9UM17 Q9UM17 homo sapien
19	120	22.0	1171	11	054893 054893 mus musculus

ALIGNMENTS

20	118.5	21.7	1256	4	Q75085	Q75085 homo sapien
21	117.5	21.6	596	4	Q60705	Q60705 homo sapien
22	116.5	21.4	400	5	Q9V7F2	Q9V7F2 drosophila
23	115.5	21.2	195	5	Q9VB77	Q9VB77 drosophila
24	115.5	21.2	197	4	Q9NUP9	Q9NUP9 homo sapien
25	115.5	21.2	197	11	Q889S2	Q889S2 mus musculus
26	114.5	21.0	337	11	Q9OYN0	Q9OYN0 mus musculus
27	114.5	21.0	591	11	Q9OYN2	Q9OYN2 mus musculus
28	113.5	20.8	138	11	Q9D130	Q9D130 mus musculus
29	113.5	20.8	283	4	Q9Y4Z5	Q9Y4Z5 homo sapien
30	113.5	20.8	288	11	Q9R0Z1	Q9R0Z1 mus musculus
31	113.5	20.8	288	11	Q9R0Z1	Q9R0Z1 mus musculus
32	113.5	20.8	470	4	Q9Y4Z3	Q9Y4Z3 homo sapien
33	113.5	20.8	617	4	Q9Y4Z4	Q9Y4Z4 homo sapien
34	113.5	20.8	661	11	Q9JKS3	Q9JKS3 mus musculus
35	113.5	20.8	723	11	Q9KXS4	Q9KXS4 mus musculus
36	113.5	20.8	723	11	Q9VWH2	Q9VWH2 mus musculus
37	113.5	20.8	734	4	Q75112	Q75112 homo sapien
38	112.5	20.6	214	11	Q9CRA2	Q9CRA2 mus musculus
39	112.5	20.6	219	11	Q9Z2S0	Q9Z2S0 rattus norv
40	112.5	20.6	233	4	Q14910	Q14910 homo sapien
41	112.5	20.6	239	11	Q9OYN1	Q9OYN1 mus musculus
42	112.5	20.6	325	4	Q9H4L9	Q9H4L9 homo sapien
43	112.5	20.6	327	11	Q9KX93	Q9KX93 mus musculus
44	112.5	20.6	329	4	Q9BPZ9	Q9BPZ9 homo sapien
45	112.5	20.6	352	6	Q9CKU1	Q9CKU1 macaca fasc

RESULT 1
ID Q9GXN0 PRELIMINARY; PRT; 1986 AA.

AC Q9GXN0;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE PDZ DOMAIN ACTIN BINDING PROTEIN SHROOM.
GN SHRM.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP MEDLINE=2005594; PubMed=10589677;
RA Hildebrand J.D., Soriano P.;
RT "Shroom, a PDZ domain-containing actin-binding protein, is required
RT for neural tube morphogenesis in mice."
RL Cell 99:485-497(1999).
DR EMBL; AF199421; AAF13269.1; -
DR HSSP; P29476; 10AV.
DR MGD; MGI:1351655; Shrm.
DR InterPro; IPR001478; PDZ.
DR Pfam; PF00595; PDZ; 1.
DR SMART; SM0228; PDZ; 1.
DR PROSITE; PS50106; PDZ; 1.
SQ SEQUENCE 1986 AA; 215261 MW; 18C5EFCB772EABF CRC64;

Query Match 75.0%; Score 409; DB 11; Length 1986;
Best Local Similarity 75.5%; Pred. No. 4.3e-32;
Matches 80; Conservative 11; Mismatches 15; Indels 0; Gaps 0;

QY 2 MRRTEDEHFRPSATLNSNTAKRYIYLEAFLEGAPMGFTLKGLHGEPLIISKVERGG 61
DB 1 MRRTEDEHFRPSATLNSNTAKRYIYLEAFLEGAPMGFTLKGLHGEPLIISKVERGG 60
QY 62 KADTLSSKLAQDEYVHINEVTLSSSRKRAVSLVKGSKYKTLRLVYR 107
DB 61 KADSVSGLAQDEYVHINEVTLSSSRKRAVSLVKGSKYKTLRLVYR 106

Fri Feb 1 08:56:32 2002

Matches 27; Conservative 13; Mismatches 10; Indels 1; Gaps 1;

RESULT 2
Q9ULL8 ID Q9ULL8 PRELIMINARY; PRT; 1502 AA.
AC Q9ULL8
DT 01-MAY-2000 (TREMREL. 13, Created)
DT 01-MAY-2000 (TREMREL. 13, Last sequence update)
DT 01-JUN-2001 (TREMREL. 17, Last annotation update)
DE KIAA1202 PROTEIN (FRAGMENT).
GN KIAA1202.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE=20039619; PubMed=10574462;
RA Nagase T., Ishikawa K., Kikuno R., Hirose M., Nomura N., Ohara O.,
RT "Prediction of the coding sequences of unidentified human genes. XV.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 6:337-345(1999).
DR EMBL; AB033028; BA86516.1; .
DR HSSP; P31016; IBE9.
DR InterPro; IPR001478; PDZ.
DR Pfam; PF00595; PDZ; 1.
DR SMART; SM00228; PDZ; 1.
DR PROSITE; PS50106; PDZ; 1.
FT NON_TER 1
SQ SEQUENCE 1502 AA; 165950 MW; DC050C9C69B78AD9 CRC64;

Query Match 48.2%; Score 262.5; DB 4; Length 1502;
Best Local Similarity 58.8%; Pred. No. 1.5e-17;
Matches 50; Conservative 17; Mismatches 17; Indels 1; Gaps 1;

QY 23 GRYIYLEAFLEGGAPWGFTLKGGLHEGPELIISKVVEGGKADTLSSKLOAGDEVVHNEV 82
DB 10 GSFQYVPVQGGAPWGFTLKGGLHEGPELTVSKIEDGGKA-ALSKMRTGDELVNINGT 68
QY 83 TLSSSRKEAVSLVKGSKYTLRLVVR 107
DB 69 PLYGSRQEAALLIKGSFRILKLIVR 93

RESULT 3
Q9H4G2 ID Q9H4G2 PRELIMINARY; PRT; 1454 AA.
AC Q9H4G2
DT 01-MAR-2001 (TREMREL. 16, Created)
DT 01-MAR-2001 (TREMREL. 16, Last sequence update)
DT 01-JUN-2001 (TREMREL. 17, Last annotation update)
DE BAI19E20.1 (KIAA1202 PROTEIN) (FRAGMENT).
GN BAI19E20.1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Wilson S.;
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL121865; CAC08472.1; .
DR InterPro; IPR001478; PDZ.
DR SMART; SM00228; PDZ; 1.
DR PROSITE; PS50106; PDZ; 1.
FT NON_TER 1
SQ SEQUENCE 1454 AA; 160615 MW; 7931571D7DA80DE8 CRC64;

Query Match 24.7%; Score 134.5; DB 4; Length 1454;
Best Local Similarity 52.9%; Pred. No. 0.0001;

Matches 27; Conservative 13; Mismatches 10; Indels 1; Gaps 1;
QY 57 VEEGKADTLSSKLOAGDEVVHNEVTLSSSRKEAVSLVKGSKYTLRLVVR 107
DB 1 IEDGGKA-ALSKMRTGDELVNINGTPLYGSRQEAALLIKGSFRILKLIVR 50

RESULT 4
Q9H4T0 ID Q9H4T0 PRELIMINARY; PRT; 39 AA.
AC Q9H4T0
DT 01-MAR-2001 (TREMREL. 16, Created)
DT 01-MAR-2001 (TREMREL. 16, Last sequence update)
DT 01-MAR-2001 (TREMREL. 16, Last annotation update)
DE BA554P16.1 (KIAA1202) (FRAGMENT).
GN BA554P16.1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Bird C.;
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL359272; CAC12948.1; .
FT NON_TER 39
SQ SEQUENCE 39 AA; 4260 MW; 2773DAE613D25D3F CRC64;

Query Match 23.5%; Score 128; DB 4; Length 39;
Best Local Similarity 67.6%; Pred. No. 4.3e-06;
Matches 23; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

QY 23 GRYIYLEAFLEGGAPWGFTLKGGLHEGPELIISK 56
DB 6 GSFQYVPVQGGAPWGFTLKGGLHEGPELTVSK 39

RESULT 5
Q9PU46 ID Q9PU46 PRELIMINARY; PRT; 298 AA.
AC Q9PU46
DT 01-MAY-2000 (TREMREL. 13, Created)
DT 01-MAY-2000 (TREMREL. 13, Last sequence update)
DT 01-JUN-2001 (TREMREL. 17, Last annotation update)
DE ALPHA-ACTININ ASSOCIATED LIM PROTEIN, SKELETAL MUSCLE ISOFORM (SKALP)
(FRAGMENT).
GN ALP.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99436131; PubMed=10506181;
RA Ponies P., Macalima T., Beckerle M.C.;
RT "Purification and characterization of an alpha-actinin-binding PDZ-LIM
RT protein that is up-regulated during muscle differentiation.";
J. Biol. Chem. 274:29242-29250(1999).
RL EMBL; AJ249219; CAB53971.1; .
DR InterPro; IPR001478; PDZ.
DR Pfam; PF00595; PDZ; 1.
DR PROSITE; PS50106; PDZ; 1.
DR SMART; SM00228; PDZ; 1.
FT NON_TER 298
SQ SEQUENCE 298 AA; 32094 MW; 27D6F24FC6141E61 CRC64;

Query Match 23.0%; Score 125.5; DB 13; Length 298;
Best Local Similarity 37.7%; Pred. No. 0.0001;
Matches 29; Conservative 16; Mismatches 29; Indels 3; Gaps 2;

DT 01-JUN-2001 (Tremblrel. 17, Created)
 DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
 DE VERTEBRATE HOMOLOGY OF C. ELEGANS LIN-7 TYPE 2.
 GN VELI2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=HIPPOCAMPUS;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
 RA Schriml L.M., Stauber F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
 RA Hayashizaki Y.
 RT "Functional annotation of a full-length mouse cDNA collection."
 RL Nature 409:685-690(2001).
 DR EMBL; AK019299; BAB31655.1;
 DR MGD; MGI:1330858; Veli2.
 DR InterPro; IPR001478; PDZ.
 DR Pfam; PF00595; PDZ; 1.
 DR SMART; SM00228; PDZ; 1.
 DR PROSITE; PS50106; PDZ; 1.
 SQ SEQUENCE 234 AA; 25732 MW; 76919D8B1C29E53E CRC64;

Query Match 22.7%; Score 123.5; DB 11; Length 234;
 Best Local Similarity 36.0%; Pred. No. 0.00012;
 Matches 36; Conservative 20; Mismatches 35; Indels 9; Gaps 4;

QY 13 ATLNNTATKG----RYYLEAFLEGGAPWFTLKGLEHGEPLIISKVEGGKADTLSS 68
 DB 77 ATVAFTASEGHAHPVVELPKTDG---LGFNIMGKQNSPIYISRVIPGVADR-HG 132
 QY 69 KIQAGDEVVHNEVTLSSSRKE-AVSLVKGSKYKTLRLVVR 107
 DB 133 GLKRGDQLLSVNGSVGEHHEKRAVELLKAQGSVKLVVR 172

RESULT 10
 ID Q9PU47 PRELIMINARY; PRT; 315 AA.
 AC Q9PU47;
 DT 01-MAY-2000 (Tremblrel. 13, Created)
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
 DE ALPHA-ACTININ ASSOCIATED LIM PROTEIN, SMOOTH MUSCLE ISOFORM (SMALP).
 GN ALP.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OX NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99436131; PubMed=10506181;
 RA Pomies P., Macalma T., Beckerle M.C.

RT "Purification and characterization of an alpha-actinin-binding PDZ-LIM
 RT protein that is up-regulated during muscle differentiation."
 RL J. Biol. Chem. 274:29242-29250(1999).
 CC -1- SIMILARITY: CONTAINS LIM DOMAIN(S). THE LIM DOMAIN BINDS 2 ZINC
 CC IONS.

DR EMBL; AJ249218; CAB53970.1;
 DR InterPro; IPR001781; LIM.
 DR InterPro; IPR001478; PDZ.
 DR Pfam; PF00412; LIM; 1.
 DR Pfam; PF00595; PDZ; 1.
 DR ProDom; PD000094; LIM; 1.
 DR SMART; SM00132; LIM; 1.
 DR SMART; SM00228; PDZ; 1.
 DR PROSITE; PS50023; LIM_DOMAIN_2; 1.
 DR PROSITE; PS50106; PDZ; 1.
 DR LIM motif; Metal-binding; Zinc.
 SQ SEQUENCE 315 AA; 34365 MW; 86414A80AB9BCA35 CRC64;

Query Match 22.7%; Score 123.5; DB 13; Length 315;
 Best Local Similarity 39.2%; Pred. No. 0.00018;
 Matches 29; Conservative 14; Mismatches 28; Indels 3; Gaps 2;

QY 32 LEGAPMGFTLKGLEHGEPLIISKVEGGKADTLSSKLGAGDEVVHNEV-TLSSSRKE 90
 DB 7 LPGAAPMGFTLGGIDFNQPLIITRITPGSKAST--ANLCPGDIIVAINGLSTENMTHND 64
 QY 91 AVSLVKGSKYKTLRL 104
 DB 65 AQERIKAAHQLSL 78

RESULT 11
 ID 017458 PRELIMINARY; PRT; 171 AA.
 AC 017458;
 DT 01-JAN-1998 (Tremblrel. 05, Created)
 DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
 DE HYPOTHETICAL 19.0 KDA PROTEIN (FRAGMENT).
 OS Schistosoma mansoni (Blood fluke).
 OC Eukaryota; Metazoa; Platyhelminthes; Rhabditophora; Neodermata;
 OC Trematoda; Digenea; Strigeidida; Schistosomatidae; Schistosomatidae;
 OC Schistosoma.
 OX NCBI_TaxID=6183;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=PUERTO RICAN;
 RA Harrop R., Wilson R.A.;
 RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF030965; AAB86566.1;
 DR HSSP; P31016; IBFE.
 DR InterPro; IPR001478; PDZ.
 DR Pfam; PF00595; PDZ; 1.
 DR SMART; SM00228; PDZ; 1.
 DR PROSITE; PS50106; PDZ; 1.
 KW Hypothetical protein.
 FT NON_TER 1
 SQ SEQUENCE 171 AA; 19000 MW; 69685F4108598063 CRC64;

Query Match 22.6%; Score 123; DB 5; Length 171;
 Best Local Similarity 34.7%; Pred. No. 9.2e-05;
 Matches 41; Conservative 20; Mismatches 39; Indels 18; Gaps 6;

QY 4 TTEDFH----KPSATLNSN---TATKG---RYYLEAFLEGGAPWFTLKGLEHGE 50
 DB 34 TVVDINGSEEVKASAKATAKATFAAFAEGHAPRVIELPKTNEG---LGFNVGKQNS 90
 QY 51 PLIISKVEGGKADTLSSKLGAGDEVVHNEVTLSSSRKE-AVSLVKGSKYKTLRLVVR 107
 DB 91 PIYISRPXPGGVADR-HGGLKRGDQLLSVNGISVESEHRAVELLKAQGVTKLVVR 147

```

RESULT 12
ID 09UPC3 PRELIMINARY; PRT; 403 AA.
AC 09UPC3;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE ANTIGEN NY-CO-37.
GN NY-CO-38.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID:9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-COLON CANCER METASTASIS TO LUNG;
RX MEDLINE:98272252; PubMed-9610721;
RA Scandan M.J., Chen Y.T., Williamson B., Gure A.O., Stockert E.,
RT "Characterization of human colon cancer antigens recognized by
RL autologous antibodies."
DR EMBL: AF039699; AAC18048.1; -.
DR HSSP: P29476; 10AV.
DR InterPro: IPR001478; PDZ.
DR Pfam: PF00595; PDZ; 2.
DR SMART: SM00228; PDZ; 2.
DR PROSITE: PS50106; PDZ; 2.
SQ SEQUENCE 403 AA; 45838 MW; AB24FB6E1D3144FD CRC64;

Query Match 22.4%; Score 122; DB 4; Length 403;
Best Local Similarity 41.4%; Pred. No. 0.00035;
Matches 29; Conservative 17; Indels 4; Gaps 3
OY 39 GFTLKGLGHEHLEPLITSKYEEGSKATLTSSKQAGGVHINVTLS-SRKEAVLVKG 97
DB 99 GLSVRGGLLEGCGCLFISHLKGGQADSVG--LQVGDVIVINISYSISCTHEVINILIR 155
OY 98 SYKTLRLVVR 107
DB 156 TKKTYSIKVR 165

RESULT 13
OQES65 PRELIMINARY; PRT; 548 AA.
ID 0QES65;
AC 0QES65;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE HARMONIN ISOFORM A1.
GN 2010016F01RIK OR USH1C.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID-10090;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE-20428180; PubMed-10973247;
RA Verpy E., Leibovici M., Zwaenepoel I., Liu X.Z., Gal A., Salem N.,
RA Mansour A., Blanchard S., Kobayashi I., Keats B.D., Silm R., Petit C.;
RT "A defect in harmonin, a PDZ domain-containing protein expressed in
RT the inner ear sensory hair cells, underlies usher syndrome type 1c."
RL Nat. Genet. 26:51-55(2000).
DR EMBL: AF228924; AAG12457.1;
DR MGD; MGI:1919338; 2010016F01RIK.
DR InterPro: IPR001478; PDZ.
DR Pfam: PF00595; PDZ; 3.
DR SMART: SM00228; PDZ; 3.
DR PROSITE: PS50106; PDZ; 3.
SQ SEQUENCE 548 AA; 61712 MW; 44A245AA32091070 CRC64;

```

```

Query Match      22.4%; Score 122; DB 11; Length 548;
Best Local Similarity 41.4%; Pred. No. 0.00051;
Matches 29; Conservative 20; Mismatches 17; Indels 4; Gaps 3;

QY 39 GFTLKGLEHGEPIITISKVEBGKADTLSSKLGADGVHINEVTLSS-SRKEAVALYVG 97
      |:::||||| |::: |:::|::: |:::|::: |:::|::: |:::|:::
Db 99 GLSVRGGLFEFGCGFLFISHLIRKGGQADSVG--LQYGEIVIRNGYSISSCTHEEVINLIR 155
      |:::|::: |::: |::: |::: |::: |::: |::: |::: |::: |::: |:::
      98 SYKTLRLVLR 107
      : |::: |:::
Db 156 TKKTIVSIKVR 165

RESULT 14
Q9UM04 PRELIMINARY; PRT; 552 AA.
ID Q9UM04;
AC Q9UM04;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DE 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DEF AUTOIMMUNE ENTEROPATHY-RELATED ANTIGEN AIE-75.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=SMALL INTESTINE;
RX MEDLINE=99431904; PubMed=10500064;
RA Kobayashi I., Imamura K., Kubota M., Ishikawa S., Yamada M.,
RA Tonoki H., Okano M., Storch W.F., Moriwuchi T., Sakiyama Y.,
RA Kobayashi K.;
RT "Identification of an autoimmune enteropathy-related 75-kilodalton
RT antigen."
RT RL Gastroenterology 117:823-830(1999).
DR EMBL: AB018687; BAA81740.1; -.
DR HSSP: P29476; LOAY.
DR InterPro: IPR001478; PDZ.
DR Pfam: PR00595; PDZ; 3.
DR SMART: SM00228; PDZ; 3.
DR PROSITE: PS0106; PDZ; 3.
SQ SEQUENCE 552 AA; 62210 MW; 7E75CE8E73C57F41 CRC64;

Query Match      22.4%; Score 122; DB 4; Length 552;
Best Local Similarity 41.4%; Pred. No. 0.00052;
Matches 29; Conservative 20; Mismatches 17; Indels 4; Gaps 3;

QY 39 GFTLKGLEHGEPIITISKVEBGKADTLSSKLGADGVHINEVTLSS-SRKEAVALYVG 97
      |:::||||| |::: |:::|::: |:::|::: |:::|::: |:::|:::
Db 99 GLSVRGGLFEFGCGFLFISHLIRKGGQADSVG--LQYGEIVIRNGYSISSCTHEEVINLIR 155
      |:::|::: |::: |::: |::: |::: |::: |::: |::: |::: |::: |:::
      98 SYKTLRLVLR 107
      : |::: |:::
Db 156 TKKTIVSIKVR 165

RESULT 15
Q9Y6N9 PRELIMINARY; PRT; 652 AA.
ID Q9Y6N9;
AC Q9Y6N9;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DE 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DEF ANTIGEN NY-CO-38.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]

```

RP SEQUENCE FROM N.A.
 RX MEDLINE-9827252; PubMed-9610721;
 RA Scanlan M.J., Chen Y.T., Williamson B., Gure A.O., Stockert E.,
 RA Gordan J.D., Tureci O., Sahin U., Pfreundschuh M., Old L.J.;
 RT "Characterization of human colon cancer antigens recognized by
 RT autologous antibodies."
 RL Int. J. Cancer 76:652-658(1998).
 DR EMBL; AF039700; AAC18049.1; -
 DR HSP; P29476; ICAV;
 DR InterPro; IPR001478; PDZ.
 DR Pfam; PF00595; PDZ; 3.
 DR SMART; SM0228; PDZ; 3.
 DR PROSITE; PS50106; PDZ; 3.
 SQ SEQUENCE 652 AA; 73336 MW; 3D7CA3BDB1174156 CRC64;

Query Match 22.4%; Score 122; DB 4; Length 652;
 Best Local Similarity 41.4%; Pred. No. 0.00064;
 Matches 29; Conservative 20; Mismatches 17; Indels 4; Gaps 3;
 Qy 39 GFTLKGLEHCEPLIISKVEGGKADTLSSKLOAGDEVVHINEVTLSS-SRKEAVSLVKG 97
 Db 99 GLSVRGGLFCGGLFISHLKGQADSVG--LQVGDEIVRINGYSISCTHEEVINLIR- 155
 Qy 98 SYKTLRLVVR 107
 Db 156 TKKTVSIKVR 165

Search completed: January 31, 2002, 21:05:35
 Job time: 98 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 20:54:52 ; Search time 56.12 Seconds
(Without alignments)
250.783 Million cell updates/sec

Title: US-09-641-831-2

Perfect score: 1029
Sequence: 1 MKRTEDFKPKSNATLNSNTA.....CTYPHOHNCRNKRAYLSR 190

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : A.Geneseq.1101.*
1: /SID52/gcgdata/geneseq/geneseqp/AA1980.DAT:*
2: /SID52/gcgdata/geneseq/geneseqp/AA1981.DAT:*
3: /SID52/gcgdata/geneseq/geneseqp/AA1982.DAT:*
4: /SID52/gcgdata/geneseq/geneseqp/AA1983.DAT:*
5: /SID52/gcgdata/geneseq/geneseqp/AA1984.DAT:*
6: /SID52/gcgdata/geneseq/geneseqp/AA1985.DAT:*
7: /SID52/gcgdata/geneseq/geneseqp/AA1986.DAT:*
8: /SID52/gcgdata/geneseq/geneseqp/AA1987.DAT:*
9: /SID52/gcgdata/geneseq/geneseqp/AA1988.DAT:*
10: /SID52/gcgdata/geneseq/geneseqp/AA1989.DAT:*
11: /SID52/gcgdata/geneseq/geneseqp/AA1990.DAT:*
12: /SID52/gcgdata/geneseq/geneseqp/AA1991.DAT:*
13: /SID52/gcgdata/geneseq/geneseqp/AA1992.DAT:*
14: /SID52/gcgdata/geneseq/geneseqp/AA1993.DAT:*
15: /SID52/gcgdata/geneseq/geneseqp/AA1994.DAT:*
16: /SID52/gcgdata/geneseq/geneseqp/AA1995.DAT:*
17: /SID52/gcgdata/geneseq/geneseqp/AA1996.DAT:*
18: /SID52/gcgdata/geneseq/geneseqp/AA1997.DAT:*
19: /SID52/gcgdata/geneseq/geneseqp/AA1998.DAT:*
20: /SID52/gcgdata/geneseq/geneseqp/AA1999.DAT:*
21: /SID52/gcgdata/geneseq/geneseqp/AA2000.DAT:*
22: /SID52/gcgdata/geneseq/geneseqp/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1029	100.0	190	22	AAU00031
2	545	53.0	108	22	AAU00032
3	541	52.6	133	22	AAU00033
4	132.5	12.9	150	21	AA03364
5	129	12.5	113	21	AA02039
6	128.5	12.5	207	22	AA03655
7	126.5	12.3	403	20	AA07095
8	126.5	12.3	406	19	AA07036
9	126.5	12.3	481	20	AA07098
10	126.5	12.3	521	20	AA07099
11	126.5	12.3	652	20	AA07094

12	126.5	12.3	652	20	AA07050	Renal cancer assoc
13	125.5	12.2	329	19	AA058396	Homo sapiens HLIM-
14	124.5	12.1	153	21	AA036475	Homo sapiens HLIM-
15	124.5	12.1	223	20	AA097845	Human LMP-3 (HLMP-
16	124.5	12.1	223	21	AA036473	Human LMP-1 amino
17	124.5	12.1	457	20	AA097844	Human bone mineral
18	124.5	12.1	457	21	AA036472	Human LIM mineral
19	124.5	11.9	316	19	AA058397	Homo sapiens HLIM-
20	122.5	11.9	316	19	AA058397	Human prostate can
21	122.5	11.9	913	18	AA012522	Protein tyrosine p
22	122.5	11.9	913	18	AA012522	Human PTP1 protein
23	122.5	11.9	913	18	AA012516	Human 5' EST seque
24	121.5	11.8	454	17	AA088804	Human insulin rece
25	121.5	11.8	469	22	AA040870	Human polypeptide
26	121.5	11.8	632	21	AA066689	Membrane-bound pro
27	121.5	11.8	632	22	AA039084	Human polypeptide
28	121.5	11.8	632	22	AA087545	Human Prol136, Ho
29	121.5	11.8	632	22	AA065212	Human Prol136, UNO
30	119.5	11.6	86	22	AA055840	PDZ encoded domain
31	119.5	11.6	86	22	AA057631	Human LIM protein
32	119.5	11.6	86	22	AA058041	Human LIM protein
33	119.5	11.6	423	21	AA036474	Human LMP-2 (HLMP-
34	119.5	11.6	457	20	AA097843	Rat bone mineralis
35	119.5	11.6	457	21	AA036471	Human protein sequ
36	119.5	11.6	197	22	AA036600	Peptide #6916 enco
37	117	11.4	37	22	AA032879	Human protein sequ
38	117	11.4	352	22	AA095452	Neuron-associated
39	116.5	11.3	632	21	AA01385	Human cell junctio
40	116	11.3	233	20	AA029978	Human cell junctio
41	116	11.3	233	22	AA07127	Human colon cancer
42	115.5	11.2	227	22	AA053276	Human colon cancer
43	115.5	11.2	227	22	AA05128	Human PDZ protein
44	115	11.2	181	22	AA064378	Human PDZ protein
45	112	10.9	817	19	AA048101	Human discs large

ALIGNMENTS

RESULT 1
ID AAU00031 standard; Protein: 190 AA.
XX AAU00031;
AC
XX 11-MAY-2001 (first entry)
DT
XX
DE Novel human protein, NHP#1.
XX
XX
KW Human; novel human protein; NHP#1; gene therapy; drug screening;
KW obesity; high blood pressure.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Misc-feature 1
FT note="Absent if the true start codon is
FT nucleotides 4-6 of sequence in AA00039, the authors
FT are uncertain which is the true start codon"
XX
XX
XX WO200114422-A1.
XX
XX
XX 01-MAR-2001.
XX
XX 18-AUG-2000; 2000MO-US22815.
XX
XX 24-AUG-1999; 99US-0150511.
XX
XX (LEXI-) LEXICON GENETICS INC.
XX
XX Turner CA, Zambrowicz B, Friedrich G, Nehls M, Sands AT;
XX
XX WPI: 2001-218430/22.

DR N-PSDB; AAS00031.

XX Novel human polynucleotides isolated from human mammary gland cDNA

PT library, encodes novel human proteins which are useful in diagnosis,

PT drug screening, clinical trial monitoring or treating behavioural

PT disorders

XX

XX Claim 2; Page 24-25; 29pp; English.

XX

XX The sequence is a novel human protein, NHP#1, which shares

CC structural motifs with human APXL protein. Nucleotide constructs

CC encoding functional NHPs are used in gene therapy approaches for the

CC modulation of NHP expression. NHP oligonucleotides can be used as

CC hybridisation probes for screening libraries and assessing NHP gene

CC expression patterns. Also, labeled NHP nucleotide probes can be used to

CC screen a human genomic library. The NHP nucleotide sequences are also

CC useful in drug screening techniques for treating symptomatic or

CC phenotypic manifestations of perturbing the normal function of NHP in the

CC body. Examples of such manifestations may include obesity and high blood

CC pressure.

XX Sequence 190 AA;

XX

Query Match 100.0%; Score 1029; DB 22; Length 190;

Best Local Similarity 100.0%; Pred. No. 1.8e-107;

Matches 190; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MMRTEDEPHKPSATLNSNTATKGRYIYLEAFLEGAPWGFTLKGLEHGPPLIISKVEEG 60

DB 1 mmrttedfhkpsatlnsntatkgriyyleaflegapwgftlkgglehgppliskveeg 60

QY 61 GKADTLSSKLQAGDEVVHINEVTLSSSRKEAVSLVKGSKYKTLRLVVRSLPPTVTSLEFD 120

DB 61 gkadtlssklqagdevvhinevtlsssrkeavslvkgskytclrlvvrslppvtvslefd 120

QY 121 PQHPQMPRTTSFSVSTADGRHESCRRPWWKWSGPRPTTAAARWPKGCIYPTQNTC 180

DB 121 pqhpmprtrtsfsvstadgrhescrrpwwkwsgprpttwaarwpkgciyptqntc 180

QY 181 RNFKRAYLSR 190

DB 181 rnfkraylsr 190

RESULT 2

AAU00032

ID AAU00032 standard; Protein; 108 AA.

XX

AC AAU00032;

XX

DT 11-MAY-2001 (first entry)

XX

XX Novel human protein, NHP#2.

DE Human; novel human protein; NHP#2; gene therapy; drug screening;

KW obesity; high blood pressure.

XX

OS Homo sapiens.

XX

XX Key Location/Qualifiers

FT Misc_feature 1

FT /note= "Absent if the true start codon is

FT nucleotides 4-6 of sequence in AAS00040, the authors

FT are uncertain which is the true start codon"

XX

XX WO200114422-A1.

XX

XX 01-MAR-2001.

XX

XX 18-AUG-2000; 2000WO-US22815.

XX

XX 24-AUG-1999; 99US-0150511.

XX

XX (LEXI-) LEXICON GENETICS INC.

XX (LEXI-) LEXICON GENETICS INC.

XX

XX Turner CA, Zambrowicz B, Friedrich G, Nehls M, Sands AT;

XX WPI; 2001-218430/22.

DR N-PSDB; AAS00040.

XX

XX Novel human polynucleotides isolated from human mammary gland cDNA

PT library, encodes novel human proteins which are useful in diagnosis,

PT drug screening, clinical trial monitoring or treating behavioural

PT disorders

XX

XX Claim 4; Page 25; 29pp; English.

XX

XX The sequence is a novel human protein, NHP#2, which shares

CC structural motifs with human APXL protein. Nucleotide constructs

CC encoding functional NHPs are used in gene therapy approaches for the

CC modulation of NHP expression. NHP oligonucleotides can be used as

CC hybridisation probes for screening libraries and assessing NHP gene

CC expression patterns. Also, labeled NHP nucleotide probes can be used to

CC screen a human genomic library. The NHP nucleotide sequences are also

CC useful in drug screening techniques for treating symptomatic or

CC phenotypic manifestations of perturbing the normal function of NHP in the

CC body. Examples of such manifestations may include obesity and high blood

CC pressure.

XX Sequence 108 AA;

XX

Query Match 53.0%; Score 545; DB 22; Length 108;

Best Local Similarity 100.0%; Pred. No. 2e-53;

Matches 108; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MMRTEDEPHKPSATLNSNTATKGRYIYLEAFLEGAPWGFTLKGLEHGPPLIISKVEEG 60

DB 1 mmrttedfhkpsatlnsntatkgriyyleaflegapwgftlkgglehgppliskveeg 60

QY 61 GKADTLSSKLQAGDEVVHINEVTLSSSRKEAVSLVKGSKYKTLRLVVRSLPPTVTSLEFD 108

DB 61 gkadtlssklqagdevvhinevtlsssrkeavslvkgskytclrlvvrslppvtvslefd 108

RESULT 3

AAU00033

ID AAU00033 standard; Protein; 133 AA.

XX

AC AAU00033;

XX

DT 11-MAY-2001 (first entry)

XX

XX Novel human protein, NHP#3.

DE Human; novel human protein; NHP#3; gene therapy; drug screening;

KW obesity; high blood pressure.

XX

OS Homo sapiens.

XX

XX Key Location/Qualifiers

FT Misc_feature 1

FT /note= "Absent if the true start codon is

FT nucleotides 4-6 of sequence in AAS00041, the authors

FT are uncertain which is the true start codon"

XX

XX WO200114422-A1.

XX

XX 01-MAR-2001.

XX

XX 18-AUG-2000; 2000WO-US22815.

XX

XX 24-AUG-1999; 99US-0150511.

XX

XX (LEXI-) LEXICON GENETICS INC.

XX PI Turner CA, Zambrowicz B, Friedrich G, Nehls M, Sands AT;
 XX DR WPI: 2001-218430/22.
 DR N-PSDB: AAS00041.
 XX PT Novel human polynucleotides isolated from human mammary gland cDNA
 PT library, encodes novel human proteins which are useful in diagnosis,
 PT drug screening, clinical trial monitoring or treating behavioural
 PT disorders
 XX PS Claim 6; Page 25-26; 29pp; English.
 XX CC The sequence is a novel human protein, NHP#3, which shares
 CC structural motifs with human APXL protein. Nucleotide constructs
 CC encoding functional NHPs are used in gene therapy approaches for the
 CC modulation of NHP expression. NHP oligonucleotides can be used as
 CC hybridisation probes for screening libraries and assessing NHP gene
 CC expression patterns. Also, labeled NHP nucleotide sequences are also
 CC screen a human genomic library. The NHP nucleotide sequences are also
 CC useful in drug screening techniques for treating symptomatic or
 CC phenotypic manifestations of perturbing the normal function of NHP in the
 CC body. Examples of such manifestations may include obesity and high blood
 CC pressure.
 CC CC
 CC CC
 CC CC
 XX Sequence 133 AA;
 SQ
 Query Match 52.6%; Score 541; DB 22; Length 133;
 Best Local Similarity 100.0%; Pred. No. 7.8e-53;
 Matches 107; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MKRTEDFKRPATLNSNTATGRITYLEAFLEGGAPWGFILKGLHEHEPLIISVVEG 60
 DB 1 mnttedfihkpatlnsatatgryyleaflegapwgfllkglheheplllskveeg 60
 QY 61 GRADTSSKLGAGDEVYHINVEYTLSSRKREAVSLVSGSKTLLVVR 107
 DB 61 gxadtlssklgagdevyhinevylsssrkxavslvsgsktllivvr 107

RESULT 4
 AAB43364
 ID AAB43364 standard; Protein; 150 AA.
 XX AAB43364;
 AC
 XX
 XX
 DT 08-FEB-2001 (first entry)
 DE Human ORFX ORF3128 polypeptide sequence SEQ ID NO:6256.
 XX
 XX Human; open reading frame; ORFX; detection; cytosolic; hepatotropic;
 XX Vulnerary; antiproliferic; antiparkinsonian; neurotropic; neuroprotective;
 XX anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
 XX immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
 XX hypotensive; dermatological; immunosuppressive; antineoplastic;
 XX antiviral; antibacterial; antifungal; antirheumatic; antihydroid;
 XX antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
 XX neurodegenerative disorder; osteoarthritis; graft vs host disease;
 XX cardiovascular disease; diabetes mellitus; hypochromidism; SCID; AIDS;
 XX cholesterol ester storage; systemic lupus erythematosus; infection;
 XX severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 XX allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 XX bone damage; cartilage damage; antinflammatory disease; coagulation;
 XX thrombosis; contraceptive.
 XX
 XX Homo sapiens.
 XX
 XX MO200058473-A2.
 PN
 XX
 XX 05-OCT-2000.
 PD
 XX
 XX 31-MAR-2000; 2000MO-US08621.
 PF

XX 31-MAR-1999; 99US-0127607.
 PR 02-APR-1999; 99US-0127636.
 PR 05-APR-1999; 99US-0127728.
 PR 30-MAR-2000; 2000US-0540763.
 XX
 PA (CURA-) CURAGEN CORP.
 XX
 XX Shinkets RA, Leach M;
 PI
 XX WPI: 2000-602362/57.
 DR N-PSDB: AAC77573.
 XX
 PT Novel nucleic acids and peptides derived from open reading frame X,
 PT useful for treating e.g. cancers, proliferative disorders,
 PT neurodegenerative disorders and cardiovascular disease
 XX
 XX Claim 11; Page 5443; 5507pp; English.
 XX
 XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
 CC sequences have activities such as: cytosolic; hepatotropic; vulnerary;
 CC antiproliferic; antiparkinsonian; neurotropic; neuroprotective;
 CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
 CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;
 CC antineoplastic; antiviral; antifungal; antirheumatic;
 CC antihydroid; and antianaemic. The sequences can be used for determining
 CC the presence of or predisposition to, or preventing or treating
 CC pathological conditions associated with an ORFX-associated disorder. The
 CC nucleic acids can be used to express ORFX proteins in gene therapy
 CC vectors. The proteins and nucleic acids may be used to treat cancers,
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,
 CC hypertension, hypochromidism, cholesterol ester storage, systemic lupus
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
 CC nocturnal haemoglobinuria, antinflammatory disease; to enhance
 CC coagulation; to inhibit thrombosis; and as a contraceptive.
 CC
 CC
 XX Sequence 150 AA;
 SQ
 Query Match 12.9%; Score 132.5; DB 21; Length 150;
 Best Local Similarity 32.6%; Pred. No. 7.9e-07;
 Matches 46; Conservative 27; Mismatches 53; Indels 15; Gaps 7;
 QY 10 KPSATLNSNTATKG---RYILEAFLEGGAPWGFILKGLHEHEPLIISVVEGKADT 65
 DB 17 kkkatvaatlaseghahprvvelpktdg---lglfimgskqnpilystslpdyvadr 73
 QY 66 LSSKLGAGDEVYHINVEYTLSSRKREAVSLVSGSKTLLVVRSSLPYVSLF--FPDQ 122
 DB 74 -hgllrrggdlsvnvsvsgeqheavellkaagsvvlivr-yprvrleamearfehm 131
 QY 123 HPGMPMPRTTSFSVSATDGR 143
 DB 132 rsar---rrqqhqsyslsleer 149

RESULT 5
 AAG02039
 ID AAG02039 standard; Protein; 113 AA.
 XX AAG02039;
 AC
 XX
 XX 06-OCT-2000 (first entry)
 DT
 XX
 XX Human secreted protein, SEQ ID NO: 6120.
 DE
 XX
 XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
 KW gene therapy; chromosome mapping.
 KW

Homo sapiens.
EP1033401-A2.
06-SEP-2000.
21-FEB-2000; 2000EP-0200610.
26-FEB-1999; 99US-0122487.
(GEST) GENSET.
Dumas Milne Edwards J, Duclert A, Giordano J;
WPI; 2000-500381/45.
N-PSDB; AAC02045.
New nucleic acid that is a 5' expressed sequence tag (5' EST) for
obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
diagnostic, forensic, gene therapy and chromosome mapping procedures -
Claim 13; SEQ ID 6120; 71pp + CD-ROM; English.
The present sequence is a polypeptide encoded by one of a large number
of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs
were prepared from total human RNAs or polyA+ RNAs derived from 30
different tissues. EST sequences usually correspond mainly to the 3'
untranslated region (UTR) of the mRNA because they are often obtained
from oligo-dT primed cDNA libraries. Such ESTs are not well suited for
isolating cDNA sequences derived from the 5' ends of mRNAs and even in
those cases where longer cDNA sequences have been obtained, the full 5'
UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'
ends and can therefore be used to obtain full length cDNAs and genomic
DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and
chromosome mapping procedures. They are used to obtain upstream
regulatory sequences and to design expression and secretion vectors.
Sequence 113 AA;
Query Match 12.5%; Score 129; DB 21; Length 113;
Best Local Similarity 35.0%; Pred. No. 1.3e-06;
Matches 36; Conservative 17; Mismatches 46; Indels 4; Gaps 3;
32 LEGAPWGTLLKGLGHEPLIISKVEGGKADTLSSKLGQDVEVHINEV-TLSSSRKE 90
7 ltpgpgwfgllgqgkdfnmpltltsrltpgska--aagqlsgdvlvaigdyntdmthle 64
91 AVSLVKGSYKTLRLVRSLSPPVTVSLEFDP-QHPQRMPPRTR 132
65 aqnkksaasynsltlqskrpipsttappvtgplpiphqk 107
RESULT 6
ID AAE03655
ID AAE03655 standard; Protein; 207 AA.
PC AAE03655;
XX 06-AUG-2001 (first entry)
XX Human extracellular matrix and cell adhesion molecule-19 (XMAP-19).
XX Human extracellular matrix and cell adhesion molecule; XMAP;
XX gene therapy; genetic disorder; adrenoleukodystrophy; leukamias;
XX

sickle cell anaemia; thalassaemia; autoimmune disorder; acquired immune deficiency syndrome; AIDS; inflammatory disorder; acquired immune deficiency syndrome; AIDS; Addison's disease; allergy; anaemia; asthma; atherosclerosis; melanoma; Crohn's disease; diabetes mellitus; atopic dermatitis; lymphoma; cancer;


```

Query Match          12.3%   Score 126.5;   DB 20;   Length 481;
Best Local Similarity 35.6%; Pred. No. 2.le-05;
Matches 31; Conservative 24; Mismatches 25; Indels 7; Gaps
4;

QY      39  GTFTLKGGLEGHEPLIIISKVEGGKADTLSSRLQAGDEVVHINEVTLS--SRKEAVSLVKG 97
         |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::
Db       99  glsvrgglegfcglfishllkqggadsvg--lqvqdeivringysissctheevinlr- 155

QY      98  SYKTRLRVLRSLSPPTVTVSLEFDPOHP 124
         ||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db      156  tkktvskivrhiq---lipvkssodep 179
```

RESULT	10	
AAAY07099		
ID	AAAY07099	standard; Protein; 521 AA.
XX	AC	
XX	AAAY07099;	
XX		
XX	02-JUL-1999	(first entry)
DT		
XX		
DE	Colon cancer associated antigen precursor sequence.	
XX		
KW	Cancer associated antigen; diagnosis; research; treatment; human;	
KW	breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;	
KW	prostate cancer.	
XX		
XX		
OS	Homo sapiens.	
XX		
PN	W09904265-A2.	
XX		
PD	28-JAN-1999.	
XX		
PF	15-JUL-1998;	98WO-US14679.
XX		
PR	22-JUN-1998;	98US-0102322.
PR	17-JUL-1997;	97US-0896164.
PR	10-OCT-1997;	97US-0061599.
PR	10-OCT-1997;	97US-0061785.
PR	10-OCT-1997;	97US-0948705.
PR	11-OCT-1997;	97GB-0021697.
XX		
PA	(LUDW-) LUDWIG INST CANCER RES.	
XX		
PI	Chen Y, Gout I, Gure A, O'Hare M, Obata Y, Old LJ;	
PI	Pfeundsich M, Sahin U, Scanlan MJ, Stockert B;	
PI	Tureci O;	
XX		
DR	WPI; 1999-132448/11.	
XX		
PT	New isolated cancer associated nucleic acids and polypeptides -	
PT	isolated using sera from cancer patients, used to develop products	
PT	for the diagnosis, monitoring or treatment of cancers	
XX		
PS	Disclosure; Page 666-667; 787pp; English.	
XX		
CC	The invention relates to a method for diagnosing a disorder characteris-	
CC	ed by expression of a human cancer associated antigen precursor coded for	
CC	a nucleic acid molecule (NAM). The method comprises: (a) contacting a	
CC	biological sample isolated from a subject with an agent that specifically	
CC	binds to the NAM, an expression product or a fragment of an expression	
CC	product complexed with an HLA molecule; and (b) determining the	
CC	determination between the agent and the NAM or the expression product as	
CC	a determination of the disorder. The products and methods can be used in	
CC	the diagnosis, monitoring, research, or treatment of conditions	
CC	characterised by the expression of various cancer associated antigens.	
CC	The invention provides nucleic acid sequences and encoded polypeptides	
CC	which are cancer associated antigen precursors expressed in human breast	
CC	cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and	
CC	lung cancer.	


```
CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
CC lung cancer.
XX
SQ Sequence 652 AA;

Query Match 12.3%; Score 126.5; DB 20; Length 652;
Best Local Similarity 35.6%; Pred. No. 3.3e-05;
Matches 31; Conservative 24; Mismatches 25; Indels 7; Gaps 4;
QY 39 GFTLKGLGEHGEPLIISKVEEGKADTSSKLQAGDEVVHINEVTLS-SRKEAVSLVKG 97
|:::||||| |:::||||| |:::||||| |:::||||| |:::||||| |:::|||||
Db 99 glsvrgglefgcglfshlkgqadvg--lqvdeivringysissctheevinlir- 155
QY 98 SVKTLRLVRSLSPPVTVSLEFDPQRM 124
|:::||||| |:::||||| |:::||||| |:::||||| |:::||||| |:::|||||
Db 156 tkrtsvskvrhig---lipvkssdpdep 179

RESULT 13
AAW58396
ID AAW58396 standard; Protein; 329 AA.
XX
AC AAW58396;
XX
DT 11-SEP-1998 (first entry)
XX
DE Homo sapiens HLIM-2 protein.
XX
KW HLIM; HLIM-2; LIM protein; prevention; treatment; disease;
KW proliferation; regeneration; cell; tissue; organ; nerve; pancreas;
KW epithelium; accidental injury; degeneration; skin grafts;
KW cancer; arteriosclerosis; abnormal cell growth.
XX
OS Homo sapiens.
XX
PN WQ9818822-A2.
XX
PD 07-MAY-1998.
XX
PF 10-OCT-1997; 97WO-US18717..
XX
PR 28-OCT-1996; 96US-0739485.
XX
PA (INCY-) INCYTE PHARM INC.
XX
PI Bandman O, Goli SK, Hillman JL;
XX
PI WPI; 1998-272137/24.
XX
DR N-PSDB; AAV30951.
XX
PT Human LIM proteins - useful for treating and diagnosing cancer and
PT other proliferative diseases
XX
PS Claim 18; Page 53-54; 94pp; English.
XX
CC The sequence is that of the human LIM protein HLIM-2.
CC The protein, optionally expressed from gene therapy vectors,
CC can be used to treat or prevent cancer or arteriosclerosis or other
CC conditions involving abnormal regulation of cell growth.
CC Also antagonists of it, antisense molecules, ribozymes or
CC triplex-forming agents can be used to induce proliferation and
CC regeneration of cells, tissues and organs (e.g. nerves, pancreas
CC or epithelium), particularly in cases of accidental injury, or
CC degeneration, especially in context of skin grafts. Probes based
CC on the gene sequence are used to detect LIM-encoding sequences by
CC hybridisation, optionally after amplification, for diagnosis or
CC monitoring of disease, or for assessing predisposition to disease.
CC They may also be used to map the corresponding genomic sequence.
CC Similar diagnostic assays are done at the protein level using
CC antibodies as immunoassay reagents. Antibodies raised against
CC HLIM can also be used therapeutically; for targeting other active
CC agents to cells that express HLIM; in drug screening and to isolate
```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 20:56:17 ; Search time 29.66 Seconds
(without alignments)
144.155 Million cell updates/sec

Title: US-09-641-831-2

Perfect score: 1029
Sequence: 1 MWRTEDEHFKPSATLNSMTA.....CIPTQHTCNFKRAYLSR 190

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PCITUS.COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	127	12.3	327	2	US-08-739-485-10 Sequence 10, Appl
2	125.5	12.2	329	2	US-08-739-485-3 Sequence 3, Appl
3	124.5	12.1	223	4	US-09-124-238A-14 Sequence 34, Appl
4	124.5	12.1	457	4	US-09-124-238A-10 Sequence 10, Appl
5	122.5	11.9	316	2	US-08-739-485-5 Sequence 5, Appl
6	121.5	11.8	330	2	US-08-739-485-11 Sequence 11, Appl
7	121.5	11.8	454	1	US-08-166-316-2 Sequence 1, Appl
8	119.5	11.6	457	4	US-09-124-238A-1 Sequence 1, Appl
9	117	11.4	75	3	US-09-100-804-21 Sequence 21, Appl
10	116	11.3	233	2	US-09-151-611-1 Sequence 1, Appl
11	116	11.3	233	4	US-09-370-102-1 Sequence 3, Appl
12	110	10.7	297	4	US-09-151-611-3 Sequence 3, Appl
13	110	10.7	297	4	US-09-370-102-3 Sequence 14, Appl
14	108	10.5	505	1	US-08-483-278-14 Sequence 14, Appl
15	108	10.5	505	1	US-09-100-804-28 Sequence 28, Appl
16	103.5	10.1	73	3	US-09-045-632-30 Sequence 30, Appl
17	103.5	9.8	498	3	US-09-045-632-36 Sequence 36, Appl
18	100.5	9.8	541	3	US-09-045-632-28 Sequence 28, Appl
19	100.5	9.8	599	3	US-09-045-632-26 Sequence 26, Appl
20	100.5	9.8	642	3	US-09-045-632-35 Sequence 35, Appl
21	100.5	9.8	818	3	US-09-045-632-25 Sequence 25, Appl
22	100.5	9.8	861	3	US-09-045-632-34 Sequence 34, Appl
23	100.5	9.8	918	3	US-09-045-632-21 Sequence 21, Appl
24	100.5	9.8	961	3	US-09-045-632-13 Sequence 13, Appl
25	100.5	9.8	1018	3	US-09-045-632-16 Sequence 16, Appl
26	100.5	9.8	1061	3	US-09-045-632-32 Sequence 32, Appl
27	100.5	9.8	1112	3	US-09-045-632-2 Sequence 2, Appl

28	100.5	9.8	1112	3	US-09-045-632-3 Sequence 3, Appl
29	95	9.2	182	3	US-09-045-632-29 Sequence 29, Appl
30	95	9.2	283	3	US-09-045-632-27 Sequence 27, Appl
31	95	9.2	502	3	US-09-045-632-24 Sequence 24, Appl
32	95	9.2	602	3	US-09-045-632-20 Sequence 20, Appl
33	95	9.2	702	3	US-09-045-632-15 Sequence 15, Appl
34	93.5	9.1	80	5	US-08-545-860D-54 Sequence 54, Appl
35	93.5	9.1	398	3	PCT-US94-04496-54 Sequence 31, Appl
36	93.5	9.1	441	3	US-09-045-632-31 Sequence 37, Appl
37	93.5	9.1	1050	3	US-09-045-632-49 Sequence 49, Appl
38	93	9.0	1050	3	US-09-045-632-50 Sequence 50, Appl
39	93	9.0	1050	3	US-09-045-632-20 Sequence 25, Appl
40	92.5	9.0	74	3	US-09-100-804-25 Sequence 34, Appl
41	92.5	9.0	74	3	US-09-100-804-34 Sequence 11, Appl
42	91	8.8	198	3	US-09-045-632-12 Sequence 11, Appl
43	91	8.8	284	3	US-09-045-632-12 Sequence 11, Appl
44	91	8.8	507	3	US-09-045-632-13 Sequence 11, Appl
45	91	8.8	604	3	US-09-045-632-14 Sequence 14, Appl

ALIGNMENTS

RESULT 1
US-08-739-485-10
Sequence 10, Application US/08739485
Patent No. 5863898
GENERAL INFORMATION:
APPLICANT: Goli, Surya K.
APPLICANT: Hallman, Jennifer L.
APPLICANT: Bandman, Olga
TITLE OF INVENTION: NOVEL HUMAN LIM PROTEINS
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: US
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08-739,485
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0142 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 327 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: Genbank
CLONE: 1020151
US-08-739-485-10
Query Match 12.3%; Score 127; DB 2; Length 327;

Db 122 GISLEDNR 129

RESULT 7

US-08-166-316-2

Sequence 2, Application US/08166316

Patent No. 5504192

GENERAL INFORMATION:

APPLICANT: Gill, Gordon N.

APPLICANT: Wu, Rui-Yun

TITLE OF INVENTION: ENDOCYTIC CODE BINDING PROTEIN

NUMBER OF SEQUENCES: 9

CORRESPONDENCE ADDRESS:

ADDRESSEE: Spensley Horn Jubas & Lubitz

STREET: 1880 Century Park East, Suite 500

CITY: Los Angeles

STATE: California

COUNTRY: USA

ZIP: 90067

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/166.316

FILING DATE: 13-DEC-1993

CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:

NAME: Wetherell, Jr., Ph.D., John R.

REGISTRATION NUMBER: 31,678

REFERENCE/DOCKET NUMBER: PD-3204

TELECOMMUNICATION INFORMATION:

TELEPHONE: (619) 455-5100

TELEFAX: (619) 455-5110

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 454 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-166-316-2

Query Match 11.8%; Score 121.5; DB 1; Length 454;

Best Local Similarity 24.4%; Pred. No. 1.7e-05;

Matches 48; Conservative 24; Mismatches 70; Indels 55; Gaps 6;

QY 32 LEGGAPWGFTLKGLEHGEPLIISKVEGGKADTLSSKLGAGDEVVHIN-EVTLSSSRKE 90

Db 8 LEGPAPWGFRLOGGKDFNVPLSISRLTPGGKA--AQSGVGDWVLSIDGENAGSLTHIE 64

QY 91 AVSLVK-----GSYKT-----LRLVVRSLSPVTV 115

Db 65 AQNKIRACGEAPGPGQGPAGSEQTAEGLRPRGPSAYTFAPSVSLNKTAQPGPGPAD 124

QY 116 SLEFPDQHPOR--MPRTRTSFSVSTADGRHWSRPPVWVWSPRPTWAAWPKGCIY 173

Db 125 SPPOQNGOPLRLVDPDASKQRLMENTED-----WRPRQASRVPSASLPT 170

QY 174 PTOHNTCRNFKRAYLSR 190

Db 171 SQAPSSCKTPDEHLKK 187

RESULT 8

US-09-124-238A-1

Sequence 1, Application US/09124238A

Patent No. 6300127

GENERAL INFORMATION:

APPLICANT: Hair, Gregory A.

APPLICANT: Boden, Scott D.

TITLE OF INVENTION: No. 6300127el Bone Mineralization Proteins, DNA, Vectors,

FILE REFERENCE: 06148.0115

CURRENT APPLICATION NUMBER: US/09/124.238A

CURRENT FILING DATE: 1998-07-29

PRIOR APPLICATION NUMBER: 60/054,219

PRIOR FILING DATE: 1997-07-30

PRIOR APPLICATION NUMBER: 60/080,407

PRIOR FILING DATE: 1998-04-02

NUMBER OF SEQ ID NOS: 36

SOFTWARE: MS Word

SEQ ID NO 1

LENGTH: 457

TYPE: PRT

ORGANISM: Rattus norvegicus

US-09-124-238A-1

Query Match 11.6%; Score 119.5; DB 4; Length 457;

Best Local Similarity 27.7%; Pred. No. 2.9e-05;

Matches 44; Conservative 18; Mismatches 60; Indels 37; Gaps 5;

QY 32 LEGGAPWGFTLKGLEHGEPLIISKVEGGKADTLSSKLGAGDEVVHIN-EVTLSSSRKE 90

Db 8 LEGPAPWGFRLOGGKDFNVPLSISRLTPGGKA--AQAGVAVGDWVLSIDGENAGSLTHIE 65

QY 91 AVSLVKGYKTLRLVVRSLSPVTVSLEFDPQH---PQMPPTRTSFSVSTADGRHWS 147

Db 66 AQNKIRACGERLSGLSRAQP-----AQSKPKALTPPADPPRYTFAPSASLNKTARPPG 120

QY 148 CRPPW-----VKWWSRPP 160

Db 121 APPPTDSALSONGOLLROLVDPDASKQRLMENTEDWRPRP 159

RESULT 9

US-09-100-804-21

Sequence 21, Application US/09100804

Patent No. 6066472

GENERAL INFORMATION:

APPLICANT: GONEZ, LEONEL JORGE

APPLICANT: SARAS, JAN

APPLICANT: CLAESON-WELSH, LENA

APPLICANT: HELDIN, CARL-HENRIK

TITLE OF INVENTION: PRIMARY STRUCTURE AND FUNCTIONAL

TITLE OF INVENTION: EXPRESSION OF NUCLEOTIDE SEQUENCES FOR NOVEL PROTEIN

TITLE OF INVENTION: TYROSINE PHOSPHATASES

NUMBER OF SEQUENCES: 34

CORRESPONDENCE ADDRESS:

ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.

STREET: 600 ATLANTIC AVENUE

CITY: BOSTON

STATE: MASSACHUSETTS

COUNTRY: USA

ZIP: 02210

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/100.804

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/596,291

FILING DATE: 09-AUG-1996

APPLICATION NUMBER: US 08/115,573

FILING DATE: 01-SEP-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/US94/09943

FILING DATE: 01-SEP-1994

ATTORNEY/AGENT INFORMATION:

NAME: GATES, EDWARD R.

REGISTRATION NUMBER: 31,616
REFERENCE/DOCKET NUMBER: LO661/7003
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-720-3500
TELEFAX: 617-720-2441
TELEX:
INFORMATION FOR SEQ. ID NO.: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 75 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-09-100-804-21

Query Match 11.4%; Score 117; DB 3; Length 75;
Best Local Similarity 38.8%; Pred. No. 3.8e-06;
Matches 26; Conservative 12; Mismatches 27; Indels 2; Gaps 2

DY 33 EGGA PWGFTLKGLEHGEPLLIKSEVEGGKADTLSSKLQAGGEVHHINEVTLS-SRKEA 91
| | | | : : : : | | : : : :
Db 2 EDGR-GFNLTGGVDKNP LVSASINPSPADTCIPKLNEDQIVLINQRDISETHDQV 60

QY	92	VSLVKS	98
		:	
Db	61	VMEIKAS	67

```

RESULT 10
US-09-151-611-1
: Sequence 1, Application us/09151611
: Patent No. 5958731
: GENERAL INFORMATION:
: APPLICANT: yue, Henry
: APPLICANT: Au-Young, Janice
: APPLICANT: Paterson, Chandra
: TITLE OF INVENTION: CELL JUNCTION PD2 PROTEIN
: FILE REFERENCE: PR-0599 US
: CURRENT APPLICATION NUMBER: US/09/151, 611
: CURRENT FILING DATE: 1998-09-11
: NUMBER OF SEQ ID NOS: 3
: SOFTWARE: PERL Program
: SEQ ID NO 1
: LENGTH: 233
: TYPE: PRT
: ORGANISM: Homo sapiens
: FEATURE: -
: OTHER INFORMATION: 1974337
US-09-151-611-1

```

Query Match 11.3%; Score 116; DB 2; Length 233;
Best Local Similarity 32.4%; Pred. No. 2.7e-05;
Matches 36; Conservative 24; Mismatches 41; Indels 10; Gaps 5

QY 13 ALNLSNTATK-----RYTLEAFLEGGAPNPFITLKGLHEHEPILISVBEGBADTLSS 68
Db 92 ATTTAAFPASGEGRSHRPVELPRTDGG---LGFNMGGEHQNSPILISILIGVAER-HG 147
QY 69 KIADGEVVIHNEVTLSSRKE-AVSLYKGSYKIRLYVRSLSPYYSL 118
Db 148 GKRGDQLLTVGVSGVEGBHEKAAVELLKAAQDSYKLVVR-YTRVLEEM 197

```

RESULT 11
US-09-370-102-1
; Sequence 1, Application US/09370102
; Patent No. 6265547
; GENERAL INFORMATION:
; APPLICANT: Yue, Henry

```

```

1  APPLICANT: Au-Young, Janice
2  APPLICANT: Patterson, Chandra
3  TITLE OF INVENTION: CELL JUNCTION PDZ PROTEIN
4  FILE REFERENCE: PF-0599 US
5  CURRENT APPLICATION NUMBER: US/09/370,102
6  CURRENT FILING DATE: 1999-08-06
7  EARLIER APPLICATION NUMBER: 09/151,611
8  EARLIER FILING DATE: 1998-09-11
9  NUMBER OF SEQ ID NOS: 3
10 SOFTWARE: PERL Program
11 SEQ ID NO 1
12     LENGTH: 233
13     TYPE: PRT
14     ORGANISM: Homo sapiens
15     FEATURE: -
16     OTHER INFORMATION: 1974337
17 US-09-370-102-1

```

Query Match	11.38	Score 116;	DB 4;	Length 233;
Best Local Similarity	32.48;	Pred. No. 2.7e-05;		
Matches 36;	Conservative 24;	Mismatches 41;	Indels 10;	Gaps 5

QY 13 ATLSNSTATKG---RYTLEAFLEGAPWFTLKGLEGEPLITSVEESGKADITLS 68
|| : || : | | : | : | : || :
Db 92 ATVAFAASESHSRPVVELPKTDEG---LGFNWWGGKEQNSPIYSRIIPGGAVER-HG 147

```
QY      69  KLAGDEVVHINEVTLLSSRKE-AVSLVKSGYKTLRLVNRSLSPPTVSL 118
      1 : : : : : : : : 1 1 1 : : : : : : : : : : : :
Db      148  GLKRGDQLLSVNGSVGEHHEKAVELLKAAKDSVKLVLR-YTPKVL EEME 197
```

```

12 RESULT
13 US-09-151-611-3
14 Sequence 3, Application US/09151611
15 Patent No. 5958731
16 GENERAL INFORMATION:
17 APPLICANT: Yoe, Henry
18 APPLICANT: Au-Young, Janice
19 APPLICANT: Patterson, Chandra
20 TITLE OF INVENTION: CELL JUNCTION PDZ PROTEIN
21 FILE REFERENCE: PF-0599 US
22 CURRENT APPLICATION NUMBER: US/09/151,611
23 CURRENT FILING DATE: 1998-09-11
24 NUMBER OF SEQ. ID NOS: 3
25 SOFTWARE: PERL Program
26 SEQ. ID NO 3
27 LENGTH: 297
28 TYPE: PRT
29 ORGANISM: Caenorhabditis elegans
30 FEATURE: -
31 OTHER INFORMATION: g1685067
32 US-09-151-611-3

```

Query Match	10.78;	Score 110;	DB 2;	Length 297;
Best Local Similarity	29.58;	Pred. No. 0.00019;		
Matches 36;	Conservative 28;	Mismatches 44;	Indels 14;	Gaps 6

```

OY      13  ATLNSTNTATK-----RYLYLEAFLEGGAPMPELTJXGHEHEPPLIKRVEGGKADPLSS 68
Db      184  ATVAAPFAAEGHAHPRIVELKPTDGG---LQFNFMGGKEQMSPIYSIRLLPGVADR-HG 239
OY      69  KIACDEYVHINEVTLLSSKREANSLVYGSKTKLRLVYRSLSPPVYSL- -FPQPHQR 126
Db      240  GIKRSDQLLVANGVNEACHEKRAADLSSAGSVKVLIRY-PLILDMERREF---RQR 295

```

QY	127	MP	128
Db	296	IP	297

RESULT 13

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

	Qy	39	GFTLKGLEHGEPILISKVEEGGKATLSKLQAQDEVVHINEVTLS--SRKEANSLVGK	97
			: :	
	Dd	99	GISIKGGRENKMPTLISKIFKGLAADQ-TEALFVGDAIIVSNGEDLSSATHDEAVQALKK	157
			: :	
	Qy	98	SYKTRLWV---RSLSP-----PVTVSLEFDQHP-QRMP-----PRRT-	133
			: :	
	Dd	158	TGKEVLVYMKREVPYFKNSAGGTSVGWDSPASPQLQPSPGPQTNLSEAKHVPL	217
			: :	
	Qy	134	-----SFVSSTADG-----RIEWSCRP-----PWK	154
			: :	
	Dd	218	KNAVSRRCPTSDPEHYLRIGCSADGDTITFLRAKDASARSWAGIAQAINALLPWK	276

```

RESULT 15
US-08-483-278-14
; Sequence 14, Application US/08483278
; Patent No. 5686073
;
; GENERAL INFORMATION:
; APPLICANT: Campbell, Kevin P. B.
; APPLICANT: Izraghimov, Oksana B.
; APPLICANT: Ervasti, James M.
; APPLICANT: Levelle, Cynthia J.
; TITLE OF INVENTION: NUCLEIC ACID
; TITLE OF INVENTION: PROTEIN
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kevin M. Farrell, P.
; STREET: P.O. Box 999
; CITY: York Harbor
; STATE: ME
; COUNTRY: USA
; ZIP: 03911
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/483,
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/123,161
; FILING DATE: 16-SEP-93
; ATTORNEY/AGENT INFORMATION:
; NAME: Farrell, Kevin M.
; REGISTRATION NUMBER: 35,505
; REFERENCE/DOCKET NUMBER: UIR#8
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (207) 363-0558
; TELEFAX: (207) 363-0528
;
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS

```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 20:59:22 ; Search time 243.53 Seconds
(without alignments)
216.626 Million cell updates/sec

Title: US-09-641-831-2

Perfect score: 1029
Sequence: 1 MRRTEDEHFKPSATLNSNTA.....CIYPTOHNTGRNFRAYLSR 190

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 3148936 seqs, 277657034 residues

Total number of hits satisfying chosen parameters: 3148936

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Pending_Patents_AA_Main: *
1: /cgn2_6/ptodata/2/paa/PCUS_COMB.pep:*
2: /cgn2_6/ptodata/2/paa/US06_COMB.pep:*
3: /cgn2_6/ptodata/2/paa/US07_COMB.pep:*
4: /cgn2_6/ptodata/2/paa/US08_COMB.pep:*
5: /cgn2_6/ptodata/2/paa/US081_COMB.pep:*
6: /cgn2_6/ptodata/2/paa/US082_COMB.pep:*
7: /cgn2_6/ptodata/2/paa/US083_COMB.pep:*
8: /cgn2_6/ptodata/2/paa/US084_COMB.pep:*
9: /cgn2_6/ptodata/2/paa/US085_COMB.pep:*
10: /cgn2_6/ptodata/2/paa/US086_COMB.pep:*
11: /cgn2_6/ptodata/2/paa/US087_COMB.pep:*
12: /cgn2_6/ptodata/2/paa/US088_COMB.pep:*
13: /cgn2_6/ptodata/2/paa/US090_COMB.pep:*
14: /cgn2_6/ptodata/2/paa/US091_COMB.pep:*
15: /cgn2_6/ptodata/2/paa/US092_COMB.pep:*
16: /cgn2_6/ptodata/2/paa/US093_COMB.pep:*
17: /cgn2_6/ptodata/2/paa/US094_COMB.pep:*
18: /cgn2_6/ptodata/2/paa/US095_COMB.pep:*
19: /cgn2_6/ptodata/2/paa/US096_COMB.pep:*
20: /cgn2_6/ptodata/2/paa/US097_COMB.pep:*
21: /cgn2_6/ptodata/2/paa/US098_COMB.pep:*
22: /cgn2_6/ptodata/2/paa/US099_COMB.pep:*
23: /cgn2_6/ptodata/2/paa/US099_COMB.pep:*
24: /cgn2_6/ptodata/2/paa/US60_COMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	290	28.2	1616	1 PCT-US01-04098A-1172	Sequence 1172, Ap
2	290	28.2	1646	1 PCT-US01-04098A-3140	Sequence 3140, Ap
3	238.5	23.2	189	21 US-09-758-472-9626	Sequence 9626, Ap
4	134	13.0	218	1 PCT-US01-01312-1236	Sequence 1236, Ap
5	134	13.0	218	21 US-09-764-868-1236	Sequence 1236, Ap
6	133	12.9	834	24 US-60-212-413-260	Sequence 260, App
7	133	12.9	1032	24 US-60-258-275-261	Sequence 261, App
8	130.5	12.7	493	24 US-60-173-464-10230	Sequence 10230, A
9	130.5	12.7	493	24 US-60-191-637-12599	Sequence 12599, A

10	130.5	12.7	493	24	US-60-191-681-9887	Sequence 9887, Ap
11	130.5	12.7	734	1	PCT-US01-14827-13022	Sequence 13022, A
12	130.5	12.7	944	24	US-60-167-217-12593	Sequence 12593, A
13	129.5	12.6	750	24	US-60-258-251-58	Sequence 58, Appl
14	129.5	12.6	981	24	US-60-248-505-946	Sequence 946, Appl
15	128.5	12.5	163	1	PCT-US01-01312-964	Sequence 964, App
16	128.5	12.5	163	21	US-09-764-868-964	Sequence 964, App
17	128.5	12.5	207	1	PCT-US00-32990-19	Sequence 19, Appl
18	128.5	12.5	207	24	US-60-172-354-19	Sequence 19, Appl
19	128.5	12.5	399	24	US-60-173-464-23192	Sequence 23192, A
20	128.5	12.5	400	24	US-60-191-637-29251	Sequence 29251, A
21	128.5	12.5	400	24	US-60-191-681-23602	Sequence 23602, A
22	127	12.3	327	16	US-09-206-645-10	Sequence 10, Appl
23	126.5	12.3	111	12	US-08-805-819-23	Sequence 17, Appl
24	126.5	12.3	406	14	US-09-502-967-159	Sequence 159, App
25	126.5	12.3	406	19	US-09-502-967-159	Sequence 601, App
26	126	12.2	195	24	US-60-167-245-601	Sequence 14024, A
27	126	12.2	195	24	US-60-173-464-14024	Sequence 17123, A
28	126	12.2	195	24	US-60-191-637-17123	Sequence 13553, A
29	126	12.2	195	24	US-60-191-681-13553	Sequence 33346, A
30	126	12.2	204	16	US-09-270-767-33346	Sequence 48563, A
31	126	12.2	204	16	US-09-270-767-48563	Sequence 191322, A
32	125.5	12.2	204	16	US-09-270-849B-191322	Sequence 17041, A
33	125.5	12.2	129	22	US-09-834-366-17041	Sequence 14514, A
34	125.5	12.2	136	22	US-60-197-873-17041	Sequence 14514, A
35	125.5	12.2	136	24	US-09-834-366-14514	Sequence 15955, A
36	125.5	12.2	136	24	US-60-197-873-15955	Sequence 15955, A
37	125.5	12.2	139	22	US-09-834-366-15955	Sequence 3059, Appl
38	125.5	12.2	139	24	US-60-197-873-15955	Sequence 40, Appl
39	125.5	12.2	329	16	US-09-206-645-3	Sequence 34, Appl
40	125.5	12.2	376	1	PCT-US01-18569-1059	Sequence 10, Appl
41	124.5	12.1	153	1	PCT-US00-11664-40	Sequence 3059, Appl
42	124.5	12.1	223	1	PCT-US00-11664-34	Sequence 34, Appl
43	124.5	12.1	223	21	US-09-721-975-34	Sequence 34, Appl
44	124.5	12.1	457	1	PCT-US00-11664-10	Sequence 10, Appl
45	124.5	12.1	457	21	US-09-721-975-10	

ALIGNMENTS

RESULT 1
PCT-US01-04098A-1172
Sequence 1172, Application PC/TUS0104098A
GENERAL INFORMATION:
APPLICANT: Hysq, Inc.
FILE OF INVENTION: Novel Nucleic Acids and Polypeptides
TITLE REFERENCE: 21272-029
CURRENT APPLICATION NUMBER: PCT/US01/04098A
CURRENT FILING DATE: 2001-02-05
PRIOR APPLICATION NUMBER: Not Yet Assigned
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: 09/728,422
PRIOR FILING DATE: 2000-11-30
PRIOR APPLICATION NUMBER: 09/693,325
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 09/663,561
PRIOR FILING DATE: 2000-09-15
PRIOR APPLICATION NUMBER: 09/654,936
PRIOR FILING DATE: 2000-09-01
PRIOR APPLICATION NUMBER: 09/620,325
PRIOR FILING DATE: 2000-07-19
PRIOR APPLICATION NUMBER: 09/598,075
PRIOR FILING DATE: 2000-06-20
PRIOR APPLICATION NUMBER: 09/560,875
PRIOR FILING DATE: 2000-04-27
PRIOR APPLICATION NUMBER: 09/496,914
PRIOR FILING DATE: 2000-02-03
NUMBER OF SEQ ID NOS: 3960
SOFTWARE: Custom
SEQ ID NO 1172
LENGTH: 1616
TYPE: PRT

```
; ORGANISM: Homo sapiens
PCT-US01-04098A-1172

Query Match      28.2%; Score 290; DB 1; Length 1616;
Best Local Similarity 49.6%; Pred. No. 6e-19;
Matches 70; Conservative 18; Mismatches 43; Indels 10; Gaps 5;

QY 13 ATLNSNTATKGRVYILEAFLEGAPWGFTLKGLEHGEPLIISKVEEGKADTLLSSKLOA 72
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 14 AEAETRAADGGRLV--EVLGGAPWGFTLKGREHGEPLVITKIEGSKAAV-DKLLA 70
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 73 GDEVHINEVTLSSSKRAVSLVSKYKTLRLVRSLS-----PPVTVSLEFDPQHPQ-R 126
   ||| | ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 71 GDEIVGINDIGLSGFQEAICLVKGSHTKLKLVKRRSELGWRPWSHWATKFSDSHPELA 130
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 127 MPPTRTFSVSADGRHEWS 147
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 131 ASPFTSTGCPSPWS-GRHHAS 150
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 2
PCT-US01-04098A-3140
; Sequence 3140, Application PC/TUS0104098A
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: Novel Nucleic Acids and Polypeptides
; FILE REFERENCE: 21272-029
; CURRENT APPLICATION NUMBER: PCT/US01/04098A
; CURRENT FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: Not yet Assigned
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: 09/728,422
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: 09/693,325
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 09/663,561
; PRIOR FILING DATE: 2000-09-15
; PRIOR APPLICATION NUMBER: 09/654,936
; PRIOR FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: 09/620,325
; PRIOR FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 09/598,075
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: 09/560,875
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 09/496,914
; PRIOR FILING DATE: 2000-02-03
; NUMBER OF SEQ ID NOS: 3960
; SOFTWARE: Custom
; SEQ ID NO 3140
; LENGTH: 1646
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US01-04098A-3140

Query Match      28.2%; Score 290; DB 1; Length 1646;
Best Local Similarity 49.6%; Pred. No. 6.2e-19;
Matches 70; Conservative 18; Mismatches 43; Indels 10; Gaps 5;

QY 13 ATLNSNTATKGRVYILEAFLEGAPWGFTLKGLEHGEPLIISKVEEGKADTLLSSKLOA 72
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 44 AEAETRAADGGRLV--EVLGGAPWGFTLKGREHGEPLVITKIEGSKAAV-DKLLA 100
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 73 GDEVHINEVTLSSSKRAVSLVSKYKTLRLVRSLS-----PPVTVSLEFDPQHPQ-R 126
   ||| | ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 101 GDEIVGINDIGLSGFQEAICLVKGSHTKLKLVKRRSELGWRPWSHWATKFSDSHPELA 160
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 127 MPPTRTFSVSADGRHEWS 147
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 161 ASPFTSTGCPSPWS-GRHHAS 180
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

; ORGANISM: Homo sapiens
PCT-US01-04098A-472
; Sequence 472, Application US/09758472
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PH001
; CURRENT APPLICATION NUMBER: US/09/758,472
; CURRENT FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: 60/179,065
; PRIOR FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: 60/180,628
; PRIOR FILING DATE: 2000-02-04
; NUMBER OF SEQ ID NOS: 9632
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9626
; LENGTH: 189
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: SITE
; LOCATION: (148)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (150)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-758-472-9626

Query Match      23.2%; Score 238.5; DB 21; Length 189;
Best Local Similarity 57.5%; Pred. No. 4.6e-15;
Matches 46; Conservative 16; Mismatches 17; Indels 1; Gaps 1;

QY 35 GAPWGFTLKGLEHGEPLIISKVEEGKADTLLSSKLOAGDEVVHINEVTLSSSKRAVSL 94
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 19 GGTWFTLKGLEHGEPLVSKIEDGGKA-ALSKMRTGDELVNGTPLYGSRQEAAIL 77
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 95 VKGSYKTLRLVRSLSPPVT 114
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 78 IKGSFRILKLIVRRNAPVS 97
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 4
PCT-US01-01312-1236
; Sequence 1236, Application PC/TUS0101312
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc., et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT232PCT
; CURRENT APPLICATION NUMBER: PCT/US01/01312
; CURRENT FILING DATE: 2001-05-09
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1510
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1236
; LENGTH: 218
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: SITE
; LOCATION: (134)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (137)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (164)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
PCT-US01-01312-1236

Query Match      13.0%; Score 134; DB 1; Length 218;
```



```
; Sequence 10230, Application US/60173464
; GENERAL INFORMATION:
; APPLICANT: Li, Peter W.D.
; TITLE OF INVENTION: ISOLATED G-PROTEIN COUPLED RECEPTORS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING GPCR PROTEINS AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL000173
; CURRENT APPLICATION NUMBER: US/60/173,464
; CURRENT FILING DATE: 1999-12-29
; NUMBER OF SEQ ID NOS: 30269
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10230
; LENGTH: 493
; TYPE: PRT
; ORGANISM: Drosophila
US-60-173-464-10230

Query Match          12.7%; Score 130.5; DB 24; Length 493;
Best Local Similarity 34.0%; Pred. No. 0.00093;
Matches 36; Conservative 16; Mismatches 39; Indels 15; Gaps 3;

QY 33 EGGAPWGFTLKGLEHGEPLIISKVEEGKADTLSSKLOAGDEVVHINEVTLS--RKEA 91
      | | : |||: ||| | | : |||: ||| | | : |||: ||| | | : |||
Db 80 ERGSTYGTVRGREGHGTGFFVSHVEHGEAHL--KGLRIGDQILRINGFRLLDDAVHKEF 137
      | | : |||: ||| | | : |||: ||| | | : |||: ||| | | : |||

QY 92 VSLVKGSKYKTLRLVRSLSPPVTVSLEFDPQHPQMPPTRTTSFSV 137
      | | : |||: ||| | | : |||: ||| | | : |||: ||| | | : |||
Db 138 IQLVAGQDR-----VTLKVRGVGMPLVRDLPEERLSWSV 171

RESULT 9
US-60-191-637-12599
; Sequence 12599, Application US/60191637
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; TITLE OF INVENTION: NUCLEIC ACID DETECTION KITS COMPRISING
; TITLE OF INVENTION: GENE SEQUENCES EXPRESSED FROM THE DROSOPHILA GENOME, AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: CL000392
; CURRENT APPLICATION NUMBER: US/60/191,637
; CURRENT FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 42660
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12599
; LENGTH: 493
; TYPE: PRT
; ORGANISM: DROSOPHILA
US-60-191-637-12599

Query Match          12.7%; Score 130.5; DB 24; Length 493;
Best Local Similarity 34.0%; Pred. No. 0.00093;
Matches 36; Conservative 16; Mismatches 39; Indels 15; Gaps 3;

QY 33 EGGAPWGFTLKGLEHGEPLIISKVEEGKADTLSSKLOAGDEVVHINEVTLS--RKEA 91
      | | : |||: ||| | | : |||: ||| | | : |||: ||| | | : |||
Db 80 ERGSTYGTVRGREGHGTGFFVSHVEHGEAHL--KGLRIGDQILRINGFRLLDDAVHKEF 137
      | | : |||: ||| | | : |||: ||| | | : |||: ||| | | : |||

QY 92 VSLVKGSKYKTLRLVRSLSPPVTVSLEFDPQHPQMPPTRTTSFSV 137
      | | : |||: ||| | | : |||: ||| | | : |||: ||| | | : |||
Db 138 IQLVAGQDR-----VTLKVRGVGMPLVRDLPEERLSWSV 171

RESULT 10
US-60-191-681-9887
; Sequence 9887, Application US/60191681
; GENERAL INFORMATION:
; APPLICANT: Li, Peter, W.D.
; TITLE OF INVENTION: ISOLATED DROSOPHILA PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING ISOLATED DROSOPHILA PROTEINS AND
; TITLE OF INVENTION: USES THEREOF.
; FILE REFERENCE: CL000390
```

```
; CURRENT APPLICATION NUMBER: US/60/191,681
; CURRENT FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 30973
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9887
; LENGTH: 493
; TYPE: PRT
; ORGANISM: DROSOPHILA
US-60-191-681-9887

Query Match          12.7%; Score 130.5; DB 24; Length 493;
Best Local Similarity 34.0%; Pred. No. 0.00093;
Matches 36; Conservative 16; Mismatches 39; Indels 15; Gaps 3;

QY 33 EGGAPWGFTLKGLEHGEPLIISKVEEGKADTLSSKLOAGDEVVHINEVTLS--RKEA 91
      | | : |||: ||| | | : |||: ||| | | : |||: ||| | | : |||
Db 80 ERGSTYGTVRGREGHGTGFFVSHVEHGEAHL--KGLRIGDQILRINGFRLLDDAVHKEF 137
      | | : |||: ||| | | : |||: ||| | | : |||: ||| | | : |||

QY 92 VSLVKGSKYKTLRLVRSLSPPVTVSLEFDPQHPQMPPTRTTSFSV 137
      | | : |||: ||| | | : |||: ||| | | : |||: ||| | | : |||
Db 138 IQLVAGQDR-----VTLKVRGVGMPLVRDLPEERLSWSV 171

RESULT 11
PCT-US01-14827-13022
; Sequence 13022, Application PC/TUS0114827
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
; FILE REFERENCE: 21272-104
; CURRENT APPLICATION NUMBER: PCT/US01/14827
; CURRENT FILING DATE: 2001-05-16
; PRIOR APPLICATION NUMBER: 09/577,408
; PRIOR FILING DATE: 2000-05-18
; NUMBER OF SEQ ID NOS: 16102
; SOFTWARE: Custom
; SEQ ID NO 13022
; LENGTH: 734
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (636)..(651)
; OTHER INFORMATION: LIM domain proteins domain identified by eMatrix, accession
; OTHER INFORMATION: number BL00478B, p-value=4.250e-12, raw score of 14.79
; NAME/KEY: DOMAIN
; LOCATION: (558)..(734)
; OTHER INFORMATION: LIM domain containing proteins domain identified by Pfam,
; OTHER INFORMATION: accession name LIM, E-value=1.4e-28, Pfam score of 106.0
PCT-US01-14827-13022

Query Match          12.7%; Score 130.5; DB 1; Length 734;
Best Local Similarity 30.2%; Pred. No. 0.0016;
Matches 42; Conservative 20; Mismatches 68; Indels 9; Gaps 3;

QY 32 LEGGAPWGFTLKGLEHGEPLIISKVEEGKADTLSSKLOAGDEVVHINEV-TLSSSRKE 90
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 14 LTGPGWGFRLGGKDFNMPLTISRITPGSKA--AQSLSQSGDLVVAIDGVNDTHTHLE 71
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 91 AVSLVKGSKYKTLRLVRSLSPPVTVSLEFDPQH-----PQMPPTRTTSFSVSADGRH 144
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 72 AQNKIKASYNLSLTQKSKRPIPISTAPPVOTPLVPVPHQKDPALDINGSLVAPSPSP 131
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 145 EWSCREPPVKKWMSPRETWA 163
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 132 EARASPGTPTGTPELRPTFS 150

RESULT 12
US-60-167-217-12593
; Sequence 12593, Application US/60167217
```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OW protein - protein search, using sw model

Run on: January 31, 2002, 20:58:12 ; Search time 40.57 Seconds

(Without alignments)
356.745 Million cell updates/sec

Title: US-09-641-831-2

Perfect score: 1029

Sequence: 1 MMRTEDEPKPSATLNSNTA.....CIPTQHNWCRNFKRAVLSR 190

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: PIR_68:**

1: p1r1:**

2: p1r2:**

3: p1r3:**

4: p1r4:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	290	28.2	1616	2 137183	gene APXL protein
2	128.5	12.5	852	2 T10811	channel associated
3	128.5	12.5	870	2 G01974	channel associated
4	128	12.4	455	2 A55050	epilgma - human
5	127	12.3	327	2 JC4385	LIM protein - rat
6	122.5	11.9	913	1 A41109	protein-tyrosine-p
7	121.5	11.8	330	2 S71828	gene r11 protein -
8	121	11.7	911	2 I56552	synapse-associated
9	120.5	11.7	1171	2 T42372	probable guanylate
10	118.5	11.5	1256	2 JEO209	brain-specific ang
11	117.5	11.4	517	2 T47134	hypothetical prote
12	116	11.3	628	2 T09458	numb-binding prote
13	116	11.3	728	2 T09457	numb-binding prote
14	115	11.2	431	2 T16191	hypothetical prote
15	113.5	11.0	904	2 I38757	homolog of Drosoph
16	113.5	11.0	926	2 I38756	homolog of Drosoph
17	113.5	11.0	1012	2 T23160	hypothetical prote
18	110.5	10.7	317	2 T27179	hypothetical prote
19	109	10.6	624	2 T19650	hypothetical prote
20	109	10.6	1026	2 T19631	hypothetical prote
21	109	10.6	1131	2 T15617	hypothetical prote
22	108.5	10.5	322	2 S19649	cysteine proteolas
23	108.5	10.5	720	2 A45436	synapse-associated
24	108	10.5	505	2 A53214	dystrophin-associ
25	107	10.4	724	2 JH0800	postsynaptic densi
26	107	10.4	767	2 T09599	postsynaptic densi
27	106.5	10.3	1112	2 T32733	AMPA glutamate rec
28	105	10.2	926	1 A41105	protein-tyrosine-p
29	104.5	10.2	503	2 I84771	syntrophin-1 - mou

30	103.5	10.1	538	2 I59291	beta1-syntrophin -
31	103.5	10.1	960	1 A39651	discs-large tumor
32	103.5	10.1	1277	2 T14152	synaptic scaffold
33	103	10.0	505	2 S62894	alpha-syntrophin -
34	102.5	10.0	2450	2 S71625	protein-tyrosine-p
35	99.5	9.7	1387	2 JC5502	G-protein signalin
36	98.5	9.6	1367	2 T13703	tama protein - fru
37	97.5	9.5	423	2 T21570	hypothetical prote
38	97.5	9.5	440	2 T21568	hypothetical prote
39	96.5	9.4	488	2 T51379	syntrophin - Pacif
40	95	9.2	723	2 T14765	hypothetical prote
41	92	8.9	390	2 T28036	hypothetical prote
42	90	8.7	413	2 S77319	hypothetical prote
43	90	8.7	2294	2 I67630	protein tyrosine p
44	90	8.7	2466	2 I67629	protein tyrosine p
45	88	8.6	2490	1 A54971	protein-tyrosine-p

ALIGNMENTS

RESULT 1
137183
gene APXL protein - human
C:Species: Homo sapiens (man)
C:Date: 29-May-1998 #sequence-revision 29-May-1998 #text-change 21-Jul-2000
C:Accession: 137183
R:Schlaflino, M.V.; Bassi, M.T.; Rugaril, E.I.; Rennerl, A.; Galli, L.; Ballabio, A.
Hum. Mol. Genet. 4, 373-382, 1995
A:Title: Cloning of a human homologue of the Xenopus laevis APX gene from the ocular
A:Reference number: 137183; MUID:95315933
A:Accession: 137183
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1616 <RES>
A:Cross-references: EMBL:X83543; NID:9790999; PIDD:CA58534.1; PID:g1181628
C:Genetics:
A:Gene: GDB:APXL
A:Cross-references: GDB:582527
A:Map position: XP22.3-XP22.3
C:Superfamily: GIGF domain homology
F:32-104/Domain: GIGF domain homology <GIG>

Query Match 28.2%; Score 290; DB 2; Length 1616;
Best Local Similarity 49.6%; Pred. No. 4.2e-19;
Matches 70; Conservative 18; Mismatches 43; Indels 10; Gaps 5;

QY 13 ATLSNNTKRGRIYLEAFLEGAPMGFTLKGLEHGPPLIISKVEEGKADTLSSKIQ 72
DB 14 AEAEFRADGRLV--EVLGGAPMGFTLKGREHGPVITKIEEGSKAAV-DKLLA 70
QY 73 GDEVHINEVTLSSSRKDAVSLVKSRYTLRLVRSLS-----PPTVSLFDPDQPO-R 126
DB 71 GDEIVGINDLGSRGDAICLVKSHKTLKLVYKRRELBGRPRSMATATFSDHPBLA 130
QY 127 MPRTRISFVSVDGGRHWS 147
DB 131 ASPFTSTGCPSSWS-GRHNAS 150
RESULT 2
T10811
channel associated protein of synapse 2 - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 16-Jul-1999 #sequence-revision 16-Jul-1999 #text-change 21-Jan-2000
C:Accession: T10811
R:Itte, M.; Hata, Y.; Takai, Y.
Submitted to the EMBL Data Library, April 1996
A:Description: Cloning of new isoforms of PSD-95/SAP90 related genes.
A:Reference number: 217166
A:Accession: T10811
A:Status: preliminary; translated from GB/EMBL/DBJ

```
Query Match      12.3%; Score 127; DB 2; Length 327;
Best Local Similarity 32.6%; Pred. No. 0.00019;
Matches 43; Conservative 16; Mismatches 47; Indels 26; Gaps 5
```

QY	32	LECGAPWGFTLKGLEHGPLIISKVVEEGKADTSLKIQAGDEVVHIN-EVTLSSRRE	90
	:	: : : : : : : : : : : : : : : : : : : :	
DB	8	LQGPGWGFRLVGKKDFDEGFAISRTVPQSKA--ATANICIGDLTAIDGETTSMTMLE	65
QY	91	AVSLVKVGSYKTLLRVV-----RSLSPPTVSVLEFDQHQFQM-----PPR	130


```

Db 66 AQRKIGCDYNNMTLTVSRSEOKIMSPDVT---EGRKHPYKNMILASEPOEVLHIGSAHNN 122
QY 131 TRTSFSVSTADG 142
Db 123 SAMFTASPPAG 134

RESULT 6
A:1109
protein-tyrosine-phosphatase (EC 3.1.3.48), nonreceptor type 3 - human
N:Alternate names: PTPH1
C:Species: Homo sapiens (man)
C>Date: 27-Mar-1992 #sequence_revision 02-May-1994 #text_change 11-Jun-1999
C:Accession: A41109, 155698
R:Yang, O.; Tonks, N.K.
Proc. Natl. Acad. Sci. U.S.A. 89, 5949-5953, 1991
A:Title: Isolation of a cDNA clone encoding a human protein-tyrosine phosphatase with ho
A:Reference number: A41109, M01D:91296738
A:Accession: A41109
A:Molecule type: mRNA
A:Residues: 1-913 <YAN>
A:Cross-references: GB:M64572; NID:q179912; PIDN:AAA35647.1; PID:q179913
R:Kuta, S.; Iton, F.; Hinoda, Y.; Toyota, M.; Makiyuchi, Y.; Imai, K.; Yachi, A.
J. Gastronterol. 29, 727-732, 1994
A:Title: Expression of cytoskeletal-associated protein tyrosine phosphatase PTPH1 mRNA
A:Reference number: 155698; M01D:95179278
A:Accession: 155698
A>Status: translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 899-913 <RES>
A:Cross-references: GB:S76309; NID:q913165; PIDN:AAB33583.1; PID:q913166
C:Genetics:
A:Gene: GDB:PPN3
A:Cross-references: GDB:131386; OMIM:176877
A:Map position: 9q31-q931
C:Superfamily: protein-tyrosine-phosphatase, nonreceptor type 3; GIGF domain homology; F
C:Keywords: phosphoprotein, phosphoric monoester hydrolase; tyrosine-specific phosphatase
F:31-308/Domain: protein 4.1 membrane-binding domain homology <B41>
F:516-890/Domain: GIGF domain homology <GLG>
F:670-890/Domain: protein-tyrosine-phosphatase homology <PTP>
F:842/Active site: Cys (phosphocysteine intermediate) #status predicted
F:848/Binding site: substrate phosphate (Arg) #status predicted

Query Match 11.9%; Score 122.5; DB 1; Length 913;
Best Local Similarity 27.7%; Pred. No. 0.002;
Matches 33; Conservative 23; Mismatches 44; Indels 19; Gaps 4;

QY 6 EDFRPSATLNSNTATKGRY-----IYLEAFLEGGAPWFTLKGLEHGEPLI 53
Db 479 DDFRR--VTKGSTEDASQYCKDNKNDSDYLIRITPDDKGFGLKGGVQDKMPLV 536

QY 54 ISKYEKGADTLSSKIQAGDEVVHINEVTLSS-SRKEAVSLVSGSYKT-----LRLVLR 107
Db 537 VSRINPSPADTCIPKINEGDIIVLINGRDISEHTHQDVVAFIRASRSHSRELALVIR 595

RESULT 7
A:71828
gene r11 protein - rat
N:Alternate names: LIM-domain protein r11
C:Species: Rattus norvegicus (Norway rat)
C>Date: 06-Dec-1996 #sequence_revision 06-Dec-1996 #text_change 05-Nov-1999
C:Accession: S71828, S39246
R:Schaefer, R.
Submitted to the EMBL Data Library, March 1996
A:Reference number: S71828
A:Accession: S71828
A:Molecule type: mRNA
A:Residues: 1-330 <SCH>
A:Cross-references: EMBL:X76454; NID:q755780; PIDN:CAA53992.1; PID:q887580
C:Genetics:

```

```

A:Gene: r11
C:Superfamily: LIM metal-binding repeat homology; GIGF domain homology
F:8-80/Domain: GIGF domain homology <GLG>
F:255-305/Domain: LIM metal-binding repeat homology <LIM2>

Query Match 11.8%; Score 121.5; DB 2; Length 330;
Best Local Similarity 32.0%; Pred. No. 0.00065;
Matches 41; Conservative 15; Mismatches 51; Indels 21; Gaps 5;

QY 32 LFGGAPWFTLKGLEHGEPLIISKYEKGADTLSSKIQAGDEVVHIN-EVTLSSSRE 90
Db 7 LRGPSPWGFVLVGRFSAPLTIRVHAGSKA--ALALCPGDSIQAINGESTELMTILE 64

QY 91 AVSLVKSYSKTLRLVVRSGSPYTVS-----LEFPPQHPQMPRTKTS--F 135
Db 65 AQRKIGCDHDLTL---SVSRPENKNWPSSPNDKQAHRHIDPEADQSPATSRSSIS 121

QY 136 SVSTADGR 143
Db 122 GISELDNR 129

RESULT 8
A:156552
synapse-associated protein 97 - rat
C:Species: Rattus norvegicus (Norway rat)
C>Date: 26-Jul-1996 #sequence_revision 26-Jul-1996 #text_change 21-Jan-2000
C:Accession: 156552
R:Miller, B.M.; Kistner, U.; VEH, R.W.; Cases-Langhoff, C.; Becker, B.; Gundelfinger,
J. Neurosci. 15, 2354-2366, 1995
A:Title: Molecular characterization and spatial distribution of SAP97, a novel presyn
A:Reference number: 156552; M01D:95198112
A:Accession: 156552
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-911 <RES>
A:Cross-references: EMBL:U14950; NID:g642455; PIDN:AAA79976.1; PID:g642456
C:Superfamily: discs-large tumor suppressor; GIGF domain homology; guanylate kinase h
F:229-307/Domain: GIGF domain homology <GLG1>
F:323-401/Domain: GIGF domain homology <GLG2>
F:587-645/Domain: SH3 homology <SH3>
F:722-899/Domain: guanylate kinase homology <GKI>

Query Match 11.8%; Score 121; DB 2; Length 911;
Best Local Similarity 25.8%; Pred. No. 0.0028;
Matches 42; Conservative 30; Mismatches 55; Indels 36; Gaps 7;

QY 3 RTEDFHKPSATLNSNTATKGRYTL-EAFL-----EGGAPWFTLKGGL 46
Db 424 QSVNHNVPSSSYLTQTPASPARYSPIKAVYLDGDEITRPPKRVLLHSGTGLGNIVG- 482

QY 47 EHGEPLIISKYEKGADTLSSKIQAGDEVVHINEVTL-SSRKEAVSLVSGSYKTRLV 105
Db 483 EDGEGIFSLFLLAGPAD-LSGELRKGDRIISVSVLDRAASHQMAALAKNA----- 534

QY 106 VRSLSPPVTVSLERDPQHPQMPPR-----TRTSFSVSTADG 142
Db 535 ---GQAVTIVAYRPEYSRFEAKIHDLRETMNNSVSSGSG 573

RESULT 9
A:742372
probable guanylate kinase (EC 2.7.4.8) 1, membrane-associated, splice form b - mouse
C:Species: Mus musculus (house mouse)
C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 18-Aug-2000
C:Accession: T42372
R:Dobrosotskaya, I.; Guy, R.K.; James, G.L.
J. Biol. Chem. 272, 31589-31597, 1997
A:Title: MAGI-1: a membrane-associated guanylate kinase with a unique arrangement of
A:Reference number: 222139; M01D:98058950
A:Accession: T42372

```

A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1171 <DOB>
A:Cross-references: EMBL:AF027503; NID:g2702346; PID:g2702347; PIDN:AAB91995.1
A:Reference number: Z24375
A:Experimental source: strain C57 Black/6 x CBA
C:Genetics:
A:Gene: Magi-1
C:Superfamily: WW repeat homology
C:Keywords: alternative splicing; phosphotransferase
F:300-337/Domain: WW repeat homology <WR1>
F:347-384/Domain: WW repeat homology <WR2>

Query Match 11.7%; Score 120.5; DB 2; Length 1171;
Best Local Similarity 26.5%; Pred. No. 0.0044;
Matches 35; Conservative 22; Mismatches 46; Indels 29; Gaps 5;
QY 9 HKPS--ATLNSNTATKGRY-----IYLEAFLEGAPWGFTLKGGLEHGE 50
DB 998 HAPSOQGTQETRTITTKQDSQFQFKGPAQAQODFYTVLELGRGAKGFGFSLRGGRYNNM 1057
QY 51 PLIISKVEEGKADTLSSKLAQGEVHHIN-EVTLSRRKEAVSLVSGYKTLRLVVR-- 107
DB 1058 DIYVRLAEDGPAER-CGKMRIQDEILNGETTKNMKHSRAIELIKNGRRVRLFLRG 1116
QY 108 -----SLSP 112
DB 1117 DGSVPEYAMIPP 1128

RESULT 10

Brain-specific angiogenesis inhibitor-associated protein 1 - human
N:Alternate names: BAI1-associated protein 1; BAP1 [misnomer]
C:Species: Homo sapiens (man)
C:Date: 21-Aug-1998 #sequence_revision 21-Aug-1998 #text_change 18-Aug-2000
C:Accession: JF0209
R:Shiratsuchi, T.; Futamura, M.; Oda, K.; Nishimori, H.; Nakamura, Y.; Tokino, T.
Biochem. Biophys. Res. Commun. 247, 597-604, 1998
A:Title: Cloning and characterization of BAI1-associated protein 1: A PDZ domain-containing
A:Reference number: JF0209; MUID:98321173
A:Accession: JF0209
A:Molecule type: mRNA
A:Residues: 1-1256 <SHI>
A:Cross-references: GB:AB010894; NID:g3370997
A:Experimental source: brain
C:Genetics:
A:Gene: GDB:BAIAP1; BAP1
A:Cross-references: GDB:9864783
A:Map position: 3p14.1-3p14.1
C:Superfamily: WW repeat homology
F:300-337/Domain: WW repeat homology <WR1>
F:359-396/Domain: WW repeat homology <WR2>

Query Match 11.5%; Score 118.5; DB 2; Length 1256;
Best Local Similarity 30.4%; Pred. No. 0.0075;
Matches 31; Conservative 22; Mismatches 40; Indels 9; Gaps 4;
QY 7 DFHKPSATLNSNTATKGRYIYLEAFLEGAPWGFTLKGGLEHGEPLIISKVEEGKADTL 66
DB 1138 EFKAQATQEQDFYI---VELE---RCAGKGFSLRGGRYNNMDLYVRLAEDGPAER- 1189
QY 67 SSKLAQGEVHHIN-EVTLSRRKEAVSLVSGYKTLRLVVR 107
DB 1190 CGKMRIQDEILNGETTKNMKHSRAIELIKNGRRVRLFLK 1231

RESULT 11

T47134
Hypothetical protein DKF2p761i2312.1 - human
C:Species: Homo sapiens (man)
C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000

C:Accession: T47134
R:Ansorge, W.; Wirkner, U.; Mewes, H.W.; Weil, B.; Wiemann, S.
submitted to the Protein Sequence Database, March 2000
A:Reference number: Z24375
A:Accession: T47134
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-517 <AAA>
A:Cross-references: EMBL:AL161971
A:Experimental source: adult amygdala; clone DKFzp761i2312
C:Genetics:
A:Note: DKFzp761i2312.1

Query Match 11.4%; Score 117.5; DB 2; Length 517;
Best Local Similarity 23.4%; Pred. No. 0.0028;
Matches 46; Conservative 31; Mismatches 65; Indels 55; Gaps 10;
QY 6 EDFHKPSATLNSNTATKGRYIYLEAFLEGAPWGFTLKGGLEHGEPLIISKVEEGKADT 65
DB 49 EPFYSGERTVITRRQTVGGF-----GLSIKGAHEHNIPVYVSKISKEQRAE- 94
QY 66 LSSKLAQGEVHHINEVTLSSSRKEAVSLVSGYKTLRLVVRSLPVPVTVSLEFDPQHPQ 125
DB 95 LSGLLFGDAILQINGINVKCRHEEV-----QVLRNAGEEVTLVSLKRAPA 144
QY 126 --RMP-----PRTRTSFSVSTA--DGRH-----EWSRPPVWVWWSRPTWAA 164
DB 145 FLKLPLNEDCACAPSDQSSGTSPLCDGLHNVHPNNTDITLSCSS-----WPTSP--GL 197
QY 165 RWPQKGC---IYPTQHN 178
DB 198 RWEKRWCDLRLIPLLS 214

RESULT 12

T09458
numb-binding protein LNXp70 - mouse
C:Species: Mus musculus (house mouse)
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jul-2000
C:Accession: T09458
R:Dho, S.E.; Jacob, S.; Wolting, C.D.; French, M.B.; Rohrschneider, L.R.; McGlade, C.
J. Biol. Chem. 273, 9179-9187, 1998
A:Title: The mammalian Numb phosphotyrosine binding domain: Characterization of bindi
A:Reference number: Z16678; MUID:98204916
A:Accession: T09458
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-628 <DHO>
A:Cross-references: EMBL:AF034746; NID:g3041880; PIDN:AAC40076.1; PID:g3041881
A:Experimental source: Brain
C:Genetics:
A:Gene: lnx

Query Match 11.3%; Score 116; DB 2; Length 628;
Best Local Similarity 25.0%; Pred. No. 0.0051;
Matches 39; Conservative 27; Mismatches 52; Indels 38; Gaps 8;
QY 10 KPSATLNSNTATKGRYIYLEAFLEGAPWGFTLKGGLEHGE---PLIISKVEEGKADTL 66
DB 399 KPAATCHEKV-----VSVKDPSESLSGMTVGGASHREWDLPYIVISVEPGG-VISR 449
QY 67 SSKLAQGEVHHINEVTLSS--SRKEAVSLVSGYKTLRLVVRSLPVPVTVSLEFDPQHPQ 125
DB 450 DGRITKDILLNNGIELTEVSRTEAVAILKSAPSV-----VLKALEVKEQEAQ 499
QY 126 RMPTRTSFSVSTADGRH-----EWSRPPVWVW 155
DB 500 E-----DCSPAALDSNNHNVTPPGDWS--PSWVWV 526
RESULT 13

T09457
 numb-binding protein LNXp80 - mouse
 C:Species: Mus musculus (house mouse)
 C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 02-Sep-2000
 C:Accession: T09457
 R:Bio. S.E.: Jacob, S.; Molling, C.D.; French, M.B.; Rohrschneider, L.R.; McGlade, C.J.
 J. Biol. Chem. 273, 9179-9187, 1998
 A:Title: The mammalian Numb phosphotyrosine binding domain: Characterization of binding
 A:Reference number: 216678; MUID:98204916
 A:Accession: T09457
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-728 <DHOS>
 A:Cross-references: EMBL:AF034745; NID:g3041878; PIDN:AAAC0075.1; PID:g3041879
 A:Experimental source: developmental stage: embryo
 C:Genetics:
 A:Gene: lnx
 C:Superfamily: RING finger homology
 F:41-88/Domain: RING finger homology <RRN>

Query Match 11.3%; Score 116; DB 2; Length 728;
 Best Local Similarity 25.0%; Pred. No. 0.0062;
 Matches 39; Conservative 27; Mismatches 52; Indels 38; Gaps 8;
 QY 10 KPSATLNSMTATKGRITYEAFLGSGAPWGFTLKGLEHGE--PLISKVEEGKADTL 66
 DB 499 KPAATCTHEKV-----VSWKDPSESLGNTVGGASHENDPLIVISVEPGG-VISR 549
 QY 67 SSKLAGDEVHINEVTLSS-SRKEAVSLVKGSKYKTLRLVRSLSPPVVSLEFDPQHPO 125
 DB 550 DSRKTKGDLILNANGELTEVSRTEVALIKSPSSV-----VLALEVKQENQ 599
 QY 126 RMPPTRTISFSVSTADGRH-----EMSCRPPVWK 155
 DB 600 E-----DCSPALDLSNHNVTTPGDM--PSWVM 626

RESULT 14
 T16191
 hypothetical protein F27D9.8 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
 C:Accession: T16191
 R:Bentley, D.
 submitted to the EMBL Data Library, February 1996
 A:Description: The sequence of C. elegans cosmid F27D9.
 A:Reference number: 218473
 A:Accession: T16191
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-431 <BEN>
 A:Cross-references: EMBL:U49629; NID:g1203924; PID:g1203931; PIDN:AAA9388.1; CESP:F27D9
 C:Genetics:
 A:Gene: CESP:F27D9.8
 A:introns: 8/1; 52/3; 74/1; 96/1; 128/2; 149/3; 199/2; 255/3; 297/3; 319/3; 350/3; 390/3

Query Match 11.2%; Score 115; DB 2; Length 431;
 Best Local Similarity 27.2%; Pred. No. 0.0039;
 Matches 43; Conservative 24; Mismatches 51; Indels 40; Gaps 7;

QY 29 EAFLEEG-----APWGFTLKGLEHGE--PLISKVEEGKADTLSSKLAGDE 75
 DB 27 ETFLELATNOVVIVKPPDSGFGSLIKGSSNQMMPVLSIKIPKLPADP-CGELFTIGDA 85
 QY 76 VVHINEVTL-SSSRKEAVSLVKGSKYKTLRLVYR-----SLSPVTVSLEFDP 121
 DB 86 IYEVNIGISTEGSHDEVKMKSSGDOVTLGVNHTFHTKTPFLKPAOSLOPDGTLDDQIFDG 145
 QY 122 Q-----HPQMPPTRTISFSVSTADGRHESCRPPVWK 155
 DB 146 RSTRSHKAQSESRSLSDSKMOQHIDGREK-----KW 175

RESULT 15
 I138757
 homolog of Drosophila discs large protein, isoform 1 - human
 C:Species: Homo sapiens (man)
 C:Date: 01-Mar-1996 #sequence_revision 01-Mar-1996 #text_change 21-Jan-2000
 C:Accession: I138757
 R:Iue, R.A.; Marfatta, S.M.; Branton, D.; Chishti, A.H.
 Proc. Natl. Acad. Sci. U.S.A. 91, 9818-9822, 1994
 A:Title: Cloning and characterization of hdlg: the human homologue of the Drosophila
 A:Reference number: I138756; MUID:95024052
 A:Accession: I138757
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-904 <RES>
 A:Cross-references: EMBL:U13897; NID:g558437; PIDN:AAA50599.1; PID:g558438
 C:Genetics:
 A:Gene: GDB:DLG1
 A:Cross-references: GDB:393278; OMIM:601014
 A:Map position: 3q29-3q29
 C:Superfamily: discs-large tumor suppressor; GIGF domain homology; guanylate kinase h
 C:Keywords: alternative splicing; duplication
 F:229-307/Domain: GIGF domain homology <GIG1>
 F:324-402/Domain: GIGF domain homology <GIG2>
 F:588-646/Domain: SH3 homology <SH3>
 F:715-892/Domain: guanylate kinase homology <GKI>

Query Match 11.0%; Score 113.5; DB 2; Length 904;
 Best Local Similarity 26.2%; Pred. No. 0.014;
 Matches 42; Conservative 27; Mismatches 54; Indels 37; Gaps 8;
 QY 7 DFH-KPSATLNSMTATKGRITYL-EAFL-----EGGAPWGFTLKGLEHG 49
 DB 428 DNHVSPSSFLGQTPASPARSPYSKAVLGDDETTRERPKVYLHKGSTGIGFNIVGG-EDG 486
 QY 50 EPLISKVEEGKADTLSSKLAGDEVHINEVTL-SSSRKEAVSLVKGSKYKTLRLVNS 108
 DB 487 EGFIISFTLAGPAD-LSGELRKGDRTISVSVDLRAASHQAAALAKNA----- 535
 QY 109 LSPVTVSLEFDPQHPOQMPPTRT-----TSFSVSTADG 142
 DB 536 -GQAVTVIAQYRPEYSRPEAKIHDLRQNMNNSISGSG 574

Search completed: January 31, 2002, 20:59:16
 Job time: 64 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 21:05:37 ; Search time 26.99 Seconds
(without alignments)
258,107 Million cell updates/sec

Title: US-09-641-831-2

Perfect score: 1029
Sequence: 1 MMRTEDFHKPSATLNSNTA.....CIYPOHNTCRNFKAYLR 190

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	290	28.2	1616	APYL_HUMAN	013796 homo sapien
2	134	13.0	330	RIL_CHICK	099672 gallus gall
3	128.5	12.5	852	DLG2_RAT	063622 rattus norv
4	128.5	12.5	870	DLG2_HUMAN	015700 homo sapien
5	127	12.3	327	PDL1_RAT	P52944 rattus norv
6	122.5	11.9	913	PTN3_HUMAN	P26045 homo sapien
7	121.5	11.8	330	RIL_RAT	P36202 rattus norv
8	121	11.8	911	DLG1_RAT	P62696 rattus norv
9	119.5	11.6	849	DLG3_MOUSE	P70175 mus musculu
10	119.5	11.6	849	DLG3_RAT	062936 rattus norv
11	117.5	11.4	329	PDL1_HUMAN	000151 homo sapien
12	113.5	11.0	904	DLG1_HUMAN	012959 homo sapien
13	112	10.9	817	DLG3_HUMAN	Q92796 homo sapien
14	110	10.7	328	RIL_HUMAN	P50479 homo sapien
15	109	10.6	326	PDL1_MOUSE	070400 mus musculu
16	109	10.6	1026	PTN1_MOUSE	P28191 caenorhabd1
17	108.5	10.5	322	CYS1_HUMAN	P13277 homarus ame
18	107	10.4	724	DLG4_MOUSE	062108 mus musculu
19	107	10.4	724	DLG4_RAT	P31016 rattus norv
20	107	10.4	767	DLG4_HUMAN	P78352 homo sapien
21	105	10.2	926	PTN4_HUMAN	P29074 homo sapien
22	103.5	10.1	960	DLG1_DROME	P31007 drosophila
23	99.5	9.7	1387	RGSC_RAT	008774 rattus norv
24	92.5	9.0	1447	RGSC_HUMAN	014974 homo sapien
25	90.5	8.8	643	RHOP_MOUSE	061085 mus musculu
26	90	8.7	2485	PTND_HUMAN	012923 homo sapien
27	87.5	8.5	323	CYS2_HUMAN	P25782 homarus ame
28	87.5	8.5	933	ZO3_HUMAN	095049 homo sapien
29	86	8.4	888	ZO3_CANPA	062683 canis fami1
30	84	8.2	1435	NOS1_RABIT	019132 oryctolagus
31	83	8.0	466	EM55_MOUSE	P70290 mus musculu
32	82.5	8.0	1010	SCA4_RICPA	094175 rickettsia
33	81.5	7.9	1174	ZO2_CANPA	095168 canis fami1

34	81	7.9	1319	1	MN1_HUMAN	Q10571 homo sapien
35	80	7.8	192	1	YLB7_CAEEL	P46581 caenorhabd1
36	80	7.8	585	1	MP3_HUMAN	013368 homo sapien
37	79.5	7.7	905	1	ZO3_MOUSE	094381 mus musculu
38	79.5	7.7	1012	1	SCA4_RICSL	094380 rickettsia
39	79.5	7.7	2254	1	CCAG_RAT	054898 rattus norv
40	78.5	7.6	341	1	CATL_DROME	095029 drosophila
41	78.5	7.6	1011	1	SCA4_RICRN	094382 rickettsia
42	78.5	7.6	1013	1	SCA4_RICRH	094381 rickettsia
43	78.5	7.6	1022	1	SCA4_RICCN	052658 rickettsia
44	78	7.6	581	1	PRLR_CEREL	028235 cervus elap
45	78	7.6	616	1	PRLR_RABIT	P14787 oryctolagus

ALIGNMENTS

```

RESULT 1
APYL_HUMAN STANDARD; PRT; 1616 AA.
ID APYL_HUMAN
AC 013796;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE APICAL-LIKE PROTEIN (APYL PROTEIN).
GN APYL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Retina;
RX MEDLINE=95315933; PubMed=7795590;
RA Schiaffino V.M., Bassi M.T., Ruggeri E.I., Renieri A., Galli L.,
RA Ballabio A.;
RT "Cloning of a human homologue of the Xenopus laevis APX gene from the
RT ocular albino type 1 critical region."
RT Hum. Mol. Genet. 4:373-382(1995).
RN [2]
RP SEQUENCE OF 56-1616 FROM N.A.
RA Shen Y., Gabbas R.A.;
RA Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
CC - TISSUE SPECIFICITY: ABUNDANT IN RETINA AND MELANOMA; ALSO IN
CC BRAIN, PLACENTA, LUNG, KIDNEY, PANCREAS.
CC - SIMILARITY: SOME TO XENOPUS LAEVIS APICAL PROTEIN (APX).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb.slb.ch).
CC -----
DR EMBL: X83543; CAA58534.1;
DR EMBL: AC002365; AAC32592.1;
DR HSSP: 012959; 1PDR.
DR MIM: 300103;
DR InterPro: IPR000667; IGFBP.
DR InterPro: IPR001478; PDZ.
DR Pfam: PF00219; IGFBP; 1.
DR Pfam: PF00595; PDZ; 1.
DR SMART: SMO0228; PDZ; 1.
DR PROSITE: PS50106; PDZ; 1.
DR SIMILAR 54 112
FT
FT TO MOUSE SYNTROPHIN-1 AND RABBIT
FT DYSTROPHIN-ASSOCIATED PROTEIN.
FT
FT DOMAIN 150 153
FT DOMAIN 314 320
FT DOMAIN 343 346
FT DOMAIN 1065 1068
FT POLY-PRO.
FT POLY-ALA.
FT POLY-PRO.
SQ SEQUENCE 1616 AA; 176409 MM; 752406BSBC0B60A2 CRC64;

```

```

Query Match          28.2%; Score 290; DB 1; Length 1616;
Best Local Similarity 49.6%; Pred. No. 9.5e-19;
Matches 70; Conservative 18; Mismatches 43; Indels 10; Gaps 5;

QY 13 ATLNSMTATKGRVYILEAFLEGAPWGFTLKGGLEHGEPLIISKVEGKADTLSSKLOA 72
DB 14 AEATRAADGRLV--EVQLSGAPWGFTLKGGREHGEPLVITKIEGSKAAAV-DKLLA 70

QY 73 GDEVHNEVTLSSRKEAVSLVKGYKTLRLVRSLS-----PVTVSLFEDDPHQ-P 126
DB 71 GDEIVGINDIGLQFRQEAICLVKGSHTKLKLVKRRSELGWRPHSWHATKFSDSHPELA 130

QY 127 MPPRTTSFSVSTADGRHWS 147
DB 131 ASPFTSTSGCPSWS-GRHHAS 150

RESULT 2
RIL_CHICK
ID RIL_CHICK STANDARD; PRT; 330 AA.
AC Q9PW72;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DE LIM PROTEIN RIL (REVERSION-INDUCED LIM PROTEIN).
GN RIL.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=White leghorn;
RX MEDLINE=20379359; PubMed=10918612;
RA Fu S.-L., Maha A., Vogt P.K.;
RT "Identification and characterization of genes upregulated in cells transformed by v-Jun";
RL Oncogene 19:3537-3545(2000).
CC -!- FUNCTION: NEGATIVE REGULATOR OF CELL GROWTH.
CC -!- SIMILARITY: CONTAINS 1 LIM DOMAIN. THE LIM DOMAIN BINDS 2 ZINC IONS.
CC -!- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announcement/ or send an email to license@isb-sib.ch).
CC EMBL; AF167295; AAD46655.1;
CC InterPro; IPR001781; LIM.
CC InterPro; IPR001478; PDZ.
CC Pfam; PF00412; LIM; 1.
CC Pfam; PF00595; PDZ; 1.
CC ProDom; PD000094; LIM; 1.
CC SMART; SM00132; LIM; 1.
CC SMART; SM00228; PDZ; 1.
CC PROSITE; PS00478; LIM_DOMAIN_1; 1.
CC PROSITE; PS50023; LIM_DOMAIN_2; 1.
CC PROSITE; PS50106; PDZ; 1.
CC LIM domain; Metal-binding; Zinc.
KW DOMAIN
FT DOMAIN 8 84
FT DOMAIN 255 305
SQ SEQUENCE 330 AA; 35838 MW; 2C0E0B404301E404 CRC64;

Query Match          13.0%; Score 134; DB 1; Length 330;
Best Local Similarity 29.4%; Pred. No. 3.7e-05;
Matches 45; Conservative 17; Mismatches 43; Indels 48; Gaps 6;

```

```

QY 32 LEGGAPWGFTLKGGLEHGEPLIISKVEGKADTLSSKLOAGDEVVHIN-EVTLSSSRKE 90
DB 7 LEGPSFWGFRVLGGKDFSTPLTISRINFGSKA--ALANLCPGDIILAINGESTEAMTHLE 64
QY 91 AVSLVKGYKTLRLVV-----RSLSPV-----TVSLEFDP----- 121
DB 65 AQNKTKACVEQLLSVSRAEERSWSPFILEDKGAQAYRINIEPEPDNGPVGKRPMPHA 124
QY 122 -----OHPQMPRTTSTFSVSTA 140
DB 125 AGGSPVDSRPALSLQHPQ--PSRPHASSADAA 155

RESULT 3
DLG2_RAT
ID DLG2_RAT STANDARD; PRT; 852 AA.
AC Q63622; Q62939; P70348;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE CHANNEL ASSOCIATED PROTEIN OF SYNAPSE-110 (CHAPSIN-110) (SYNAPTIC DENSITY PROTEIN PSD-93) (DISCS, LARGE HOMOLOG 2).
GN DLG2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=963110881; PubMed=8755482;
RA Kim E., Cho K.O., Rothschild A., Sheng M.;
RT "Heteromultimerization and NMDA receptor-clustering activity of Chapsyn-110, a member of the PSD-95 family of proteins.";
RL Neuron 17:103-113(1996).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=96193770; PubMed=8625413;
RA Brenman J.E., Chao D.S., Gee S.H., McGehee A.W., Craven S.E., Santillano D.R., Wu Z., Huang F., Xia H., Peters M.F., Froehner S.C., Bret D.S.;
RA "Interaction of nitric oxide synthase with the postsynaptic density protein PSD-95 and alpha1-syntrophin mediated by PDZ domains.";
RL Cell 84:757-767(1996).
RN [3]
RP SEQUENCE FROM N.A.
RA Irie M., Hata Y., Takai Y.;
RA Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: INTERACTS WITH THE CYTOPLASMIC TAIL OF NMDA RECEPTOR SUBUNITS AS WELL AS POTASSIUM CHANNELS.
CC -!- SIMILARITY: CONTAINS 3 PDZ/DHR DOMAINS.
CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -!- SIMILARITY: BELONGS TO A GUANYLATE KINASE-LIKE DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announcement/ or send an email to license@isb-sib.ch).
CC EMBL; U49049; AAB53243.1;
CC EMBL; U50717; AAC52643.1;
CC EMBL; U53368; AAB48562.1;
CC HSP; Q12959; LPDR.
CC InterPro; IPR000619; Guanylate_kin.
CC InterPro; IPR001478; PDZ.
CC InterPro; IPR001452; SH3.
CC Pfam; PF00625; Guanylate_kin; 1.
CC Pfam; PF00595; PDZ; 3.
CC Pfam; PF00018; SH3; 1.

```

```

DR SMART; SM00072; GUKC; 1.
DR SMART; SM00228; PDZ; 3.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS00856; GUANYLATE_KINASE_1; 1.
DR PROSITE; PS50052; GUANYLATE_KINASE_2; 1.
DR PROSITE; PS50106; PDZ; 3.
DR PROSITE; PS50002; SH3; 1.
KW SH3 domain; Repeat.
FT DOMAIN 98 184
FT 279 PDZ 1.
FT 421 PDZ 2.
FT 501 PDZ 3.
FT 536 SH3.
FT 606 SH3.
FT DOMAIN 622 882
FT 181 GUANYLATE_KINASE.
FT 228 VR -> IL (IN REF. 2).
FT 228 I -> M (IN REF. 2).
FT 326 R -> K (IN REF. 2).
FT 339 D -> E (IN REF. 3).
FT 450 MISSING (IN REF. 2).
FT 454 MISSING (IN REF. 2).
FT 455 CD -> RK (IN REF. 2).
FT 464 D -> H (IN REF. 2).
FT 474 R -> P (IN REF. 2).
FT 476 A -> D (IN REF. 2).
FT 478 AAA -> LP (IN REF. 2).
FT 484 A -> S (IN REF. 2).
FT 506 H -> N (IN REF. 2).
FT 569 L -> Q (IN REF. 2).
FT 569 MISSING (IN REF. 2).
FT 641 MISSING (IN REF. 2).
FT 626 K -> A (IN REF. 3).
FT 639 F -> L (IN REF. 1).
FT 726 N -> Y (IN REF. 2).
FT 733 E -> V (IN REF. 1).
FT 749 L -> H (IN REF. 2).
FT 756 KR -> NG (IN REF. 2).
FT 792 T -> M (IN REF. 2).
FT 794 T -> M (IN REF. 2).
SQ SEQUENCE 852 AA; 94934 MW; F8DA14A8B9CF5B09 CRC64;

```

```

Query Match 12.5%; Score 128.5; DB 1; Length 852;
Best Local Similarity 28.2%; Pred. No. 0.0004; Indels 25; Gaps 7;
Matches 40; Conservative 30; Mismatches 47;

```

```

QY 10 KPSATLNSNTATKG--RYIYLEAFLEGAPWGFTLKGLEGEPLIISKVEEGKADTL 67
DB 404 QPSVTLRAVSLDEGEPRKVLH---KSTGLGFINVIG-EDGEGLFVFIILAGPAD-LS 458
QY 68 SKIAGDEVVHINEVTL-SSSRKEAVSLVKSQYKTLRLVNSLSPPVVSLEFPDQFOR 126
DB 459 GELRGDQILSVNGIDLRGASHEQAAALAKGAGOT-----VTIIAYOPEDYAR 507
QY 127 MPPRTR-----TSFSVSTADG 142
DB 508 FEAKIHDLREQMNMHSMSSGSG 529

RESULT 4
DLG2_HUMAN STANDARD; PRT; 870 AA.
AC Q15700;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE CHANNEL ASSOCIATED PROTEIN OF SYNAPSE-110 (CHAPS-110) (DISCS, LARGE
DE HOMOLOG 2).
GN DLG2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96310881; PubMed=8755482;
RA Kim E., Cho K.-O., Rothschild A., Sheng M.;
RT "Heteromultimerization and NMDA receptor-clustering activity of

```

```

RT Chapsyn-110, a member of the PSD-95 family of proteins."
RT Neuron 17:103-113(1996).
CC -1- FUNCTION: INTERACTS WITH THE CYTOPLASMIC TAIL OF NMDA RECEPTOR
CC SUBUNITS AS WELL AS POTASSIUM CHANNELS.
CC -1- SIMILARITY: CONTAINS 3 PDZ/DHR DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -1- SIMILARITY: CONTAINS A GUANYLATE_KINASE-LIKE DOMAIN.
CC -1- SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U32376; AAB04949.1; -.
CC HSSP; Q12959; IPDR.
CC MIM; 603583; -.
CC DR InterPro: IPR000619; Guanylate_kin.
CC DR InterPro: IPR001478; SH3.
CC DR Pfam; PF00625; Guanylate_kin; 1.
CC DR Pfam; PF00018; SH3; 1.
CC DR Pfam; PF00595; PDZ; 3.
CC DR SMART; SM00072; GUKC; 1.
CC DR SMART; SM00326; SH3; 1.
CC DR PROSITE; PS00856; GUANYLATE_KINASE_1; 1.
CC DR PROSITE; PS50052; GUANYLATE_KINASE_2; 1.
CC DR PROSITE; PS50106; PDZ; 3.
CC DR PROSITE; PS50002; SH3; 1.
CC SH3 domain; Repeat.
KW SH3 domain; Repeat.
FT DOMAIN 98 184
FT 279 PDZ 1.
FT 421 PDZ 2.
FT 501 PDZ 3.
FT 536 SH3.
FT 606 SH3.
FT DOMAIN 622 882
FT 181 GUANYLATE_KINASE.
FT 228 VR -> IL (IN REF. 2).
FT 228 I -> M (IN REF. 2).
FT 326 R -> K (IN REF. 2).
FT 339 D -> E (IN REF. 3).
FT 450 MISSING (IN REF. 2).
FT 454 MISSING (IN REF. 2).
FT 455 CD -> RK (IN REF. 2).
FT 464 D -> H (IN REF. 2).
FT 474 R -> P (IN REF. 2).
FT 476 A -> D (IN REF. 2).
FT 478 AAA -> LP (IN REF. 2).
FT 484 A -> S (IN REF. 2).
FT 506 H -> N (IN REF. 2).
FT 569 L -> Q (IN REF. 2).
FT 569 MISSING (IN REF. 2).
FT 641 MISSING (IN REF. 2).
FT 626 K -> A (IN REF. 3).
FT 639 F -> L (IN REF. 1).
FT 726 N -> Y (IN REF. 2).
FT 733 E -> V (IN REF. 1).
FT 749 L -> H (IN REF. 2).
FT 756 KR -> NG (IN REF. 2).
FT 792 T -> M (IN REF. 2).
FT 794 T -> M (IN REF. 2).
SQ SEQUENCE 870 AA; 97500 MW; 89C83BA0619F6F59 CRC64;

```

```

Query Match 12.5%; Score 128.5; DB 1; Length 870;
Best Local Similarity 28.2%; Pred. No. 0.0004; Indels 25; Gaps 7;
Matches 40; Conservative 30; Mismatches 47;

```

```

QY 10 KPSATLNSNTATKG--RYIYLEAFLEGAPWGFTLKGLEGEPLIISKVEEGKADTL 67
DB 404 QPSVTLRAVSLDEGEPRKVLH---KSTGLGFINVIG-EDGEGLFVFIILAGPAD-LS 458
QY 68 SKIAGDEVVHINEVTL-SSSRKEAVSLVKSQYKTLRLVNSLSPPVVSLEFPDQFOR 126
DB 459 GELRGDQILSVNGIDLRGASHEQAAALAKGAGOT-----VTIIAYOPEDYAR 507
QY 127 MPPRTR-----TSFSVSTADG 142
DB 508 FEAKIHDLREQMNMHSMSSGSG 529

RESULT 5
PDL1_RAT STANDARD; PRT; 327 AA.
AC P52944;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PDZ AND LIM DOMAIN PROTEIN 1 (LIM DOMAIN PROTEIN CLP-36) (C-TERMINAL
DE LIM DOMAIN PROTEIN 1) (ELFIN).
DE PDLIM1 OR CLIM1 OR CLP36.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;

```

```

RN  SEQUENCE FROM N.A.
RP  STRAIN=SPRAGUE-DAWLEY; TISSUE=Liver;
RX  MEDLINE=96096533; PubMed=8522188;
RA  Wang H., Harrison-Shostak D.C., Lemasters J.J., Herman B.;
RT  "Cloning of a rat cDNA encoding a novel LIM domain protein with high
RL  homology to rat RIL.";
RL  Gene 165:267-271(1995).
CC  -|- FUNCTION: CYTOSKELETAL PROTEIN THAT MAY ACT AS AN ADAPTER THAT
CC  BRINGS OTHER PROTEINS TO THE CYTOSKELETON (BY SIMILARITY).
CC  -|- SUBUNIT: INTERACTS WITH ALPHA-ACTININS, 1 2 AND 4 (BY SIMILARITY).
CC  -|- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC  -|- TISSUE SPECIFICITY: EXPRESSED MOST ABUNDANTLY IN HEART, LUNG AND
CC  LIVER, MODERATELY IN SPLEEN AND SKELETAL MUSCLE, AND AT EXTREMELY
CC  LOW LEVELS (IF AT ALL) IN TESTIS AND BRAIN TISSUES.
CC  -|- SIMILARITY: CONTAINS 1 LIM DOMAIN. THE LIM DOMAIN BINDS 2 ZINC
CC  IONS.
CC  -|- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
CC  EMBL; U23769; AAA92046.1; .
CC  InterPro; IPR001781; LIM.
CC  InterPro; IPR001478; PDZ.
CC  Pfam; PF00412; LIM; 1.
CC  Pfam; PF00595; PDZ; 1.
CC  ProDom; PD000094; LIM; 1.
CC  SMART; SM00132; LIM; 1.
CC  SMART; SM00228; PDZ; 1.
CC  PROSITE; PS00478; LIM_DOMAIN_1; FALSE_NEG.
CC  PROSITE; PS50023; LIM_DOMAIN_2; 1.
CC  PROSITE; PS50106; PDZ; 1.
CC  Cytoskeleton; LIM domain; Metal-binding; Zinc.
FT  DOMAIN 3 84
FT  DOMAIN 258 308 LIM.
FT  SEQUENCE 327 AA; 35925 MW; 4C3B5F579FBDA4D1 CRC64;
SQ  -----
Query Match 12.3%; Score 127; DB 1; Length 327;
Best Local Similarity 32.6%; Pred. NO. 0.00016;
Matches 43; Conservative 16; Mismatches 47; Indels 26; Gaps 5;
QY 32 LEGGAPWGTLKGLGHEGPEPLISKVEGGKADTLSSKLQAGDEVVHIN-EVTLSSSRKE 90
Db 8 LOGPGWGRFLVGGKDFEOPLAISRVTPGSKA--AIANLCIGDLITADIGEDTSMTHLE 65
QY 91 AVSLVKGSYKTLRLVY-----RSLSPPTVTSLEFPDQHPQRM-----PPR 130
Db 66 AQNKIKGCVDNMTLAVSRSEQINWPLVT---EEGKRHPYKKNLASEPQEVVLHIGSAHR 122
QY 131 TRTSFSVSTADG 142
Db 123 SAMPETASPAG 134
RESULT 6
PTN3_HUMAN
ID PTN3_HUMAN STANDARD; PRT; 913 AA.
AC P26045;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PROTEIN TYROSINE PHOSPHATASE, NON-RECEPTOR TYPE 3 (EC 3.1.3.48)
DE (PROTEIN-TYROSINE PHOSPHATASE H1) (PTP-H1).
GN PTPN3 OR PTPH1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
[1]
OX NCBI_TaxID=9606;
RN SEQUENCE FROM N.A.
RX MEDLINE=91296738; PubMed=1648725;
RA Yang Q., Tonks N.K.;
RT "Isolation of a cDNA clone encoding a human protein-tyrosine
RL phosphatase with homology to the cytoskeletal-associated proteins
RT band 4.1, ezrin, and talin.";
RL Proc. Natl. Acad. Sci. U.S.A. 88:5949-5953(1991).
[2]
RN SEQUENCE OF 194-896 FROM N.A.
RP TISSUE-Colon;
RX MEDLINE=9327504; PubMed=1626183;
RA Arimura Y., Hinoda Y., Itoh F., Takekawa M., Tsujisaki M., Adachi M.,
RA Imai K., Yachi A.;
RT "cDNA cloning of new protein tyrosine phosphatases in the human
RL colon.";
RL Tumour Biol. 13:180-186(1992).
[3]
RN SEQUENCE OF 899-913 FROM N.A.
RX MEDLINE=95179278; PubMed=7874267;
RA Ikuta S., Itoh F., Hinoda Y., Toyota M., Makiguchi Y., Imai K.,
RA Yachi A.;
RT "Expression of cytoskeletal-associated protein tyrosine phosphatase
RL PTPH1 mRNA in human hepatocellular carcinoma.";
RL J. Gastroenterol. 29:727-732(1994).
CC -|- FUNCTION: MAY ACT AT JUNCTIONS BETWEEN THE MEMBRANE AND THE
CC CYTOSKELETON.
CC -|- CATALYTIC ACTIVITY: PROTEIN TYROSINE PHOSPHATE + H(2)O =
CC PROTEIN TYROSINE + ORTHOPHOSPHATE.
CC -|- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -|- SIMILARITY: BELONGS TO THE NON-RECEPTOR CLASS OF THE PROTEIN-
CC TYROSINE PHOSPHATASE FAMILY.
CC -|- SIMILARITY: CONTAINS 1 BAND 4.1-LIKE DOMAIN.
CC -|- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.
CC -----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
EMBL; M64572; AAA35647.1; .
EMBL; S39392; AAB22439.2; .
EMBL; S76309; AAB33583.1; .
PIR; A41109; A41109.
HSP; P18031; 1PTT.
MIM; 176877; .
InterPro; IPR000299; Band_4.1.
InterPro; IPR001478; PDZ.
InterPro; IPR000387; TYR_phosphatase.
InterPro; IPR000242; Tyr_prot_phptase.
Pfam; PF00373; Band_4.1; 1.
Pfam; PF00595; PDZ; 1.
Pfam; PF00102; Y_phosphatase.
PRINTS; PR00700; PRTYPHPTASE.
PRINTS; PR00935; BAND41.
SMART; SM00295; B4.1; 1.
SMART; SM00228; PDZ; 1.
SMART; SM00194; PTPc; 1.
PROSITE; PS00660; BAND_4.1_1; 1.
PROSITE; PS00661; BAND_4.1_2; 1.
PROSITE; PS50057; BAND_4.1_3; 1.
PROSITE; PS50106; PDZ; 1.
PROSITE; PS00383; TYR_PHOSPHATASE_1; 1.
PROSITE; PS50056; TYR_PHOSPHATASE_2; 1.
PROSITE; PS50055; TYR_PHOSPHATASE_PTP; 1.
KW Structural protein; Cytoskeleton; Hydrolase.
FT DOMAIN 84 241 BAND 4.1-LIKE.
FT DOMAIN 510 582 PDZ.

```


FT DOMAIN 670 913 PROTEIN-TYROSINE PHOSPHATASE.
 FT ACT SITE 842 842 BY SIMILARITY.
 SO SEQUENCE 913 AA; 104029 MW; 29A539ACDE2F1515 CRC64;

Query Match 11.9%; Score 122.5; DB 1; Length 913;
 Best Local Similarity 27.7%; Pred. No. 0.0016;
 Matches 33; Conservative 23; Mismatches 44; Indels 19; Gaps 4;

QY 6 EDHFKDSATINSTATKGRY-----TYLEAFLEGAPMGFTLKGLHGEPLI 53
 DB 479 DDHHR--VTKGSTEDASQYCKRNDGSDYLVLRTPEDGKFGFNKGGVDKKRPLV 536
 QY 54 ISKVEEGKADTLSSKLGADDEVHINEVTLSS-SRREAVSLVGSYKT---LRLVVR 107
 DB 537 VSRINESPADTCIPKLNESGDYLVINGRDISEHTDQVYMFIKASRSHRELATVIR 595

RESULT 7
 ID RIL_RAT STANDARD; PRT; 330 AA.
 AC P36202;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE LIM PROTEIN RIL (REVERSION-INDUCED LIM PROTEIN) (RIT-18) (H-REV18).
 GN RIL.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 ON NCBI_TaxID-10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-SPRAGUE-DAWLEY; TISSUE-Liver;
 RX MEDLINE-95124724; Pubmed-7824279;
 RA Kless M., Scham B., Aguzzi A., Hajnal A., Klemenz R.,
 RA Schwarte-Waldhoff I., Schaefer R.,
 RT Expression of ril, a novel LIM domain gene, is down-regulated in
 RT Haas-transformed cells and restored in phenotypic revertants.
 RL Oncogene 10:61-68(1995).
 CC -1- TISSUE SPECIFICITY: DETECTED IN SEVERAL TISSUES, MOST PROMINENT IN
 CC BRAIN AND HEART OF ADULTS.
 CC -1- SIMILARITY: CONTAINS 1 LIM DOMAIN. THE LIM DOMAIN BINDS 2 ZINC
 CC IONS.
 CC -1- SIMILARITY: CONTAINS 1 PDZ/DBP DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: X76454; CAAS3992.1; -
 CC PIR: S39246; S39246.
 CC InterPro: IPR001781; LIM.
 CC InterPro: IPR001478; PDZ.
 CC Pfam: PF00412; LIM; 1.
 CC Pfam: PF00595; PDZ; 1.
 CC ProDom: PD000094; LIM; 1.
 CC SMART: SM00132; LIM; 1.
 CC SMART: SM00228; PDZ; 1.
 CC PROSITE: PS00478; LIM_DOMAIN_1; 1.
 CC PROSITE: PS50023; LIM_DOMAIN_2; 1.
 CC PROSITE: PS50106; PDZ; 1.
 CC LIM domain; Metal-binding; Zinc.
 FT DOMAIN 1 84 PDZ.
 FT DOMAIN 255 305 LIM.
 SO SEQUENCE 330 AA; 35521 MW; 84BEB1E679ADA584 CRC64;

Query Match 11.8%; Score 121.5; DB 1; Length 330;

Best Local Similarity 32.0%; Pred. No. 0.00054;
 Matches 41; Conservative 15; Mismatches 51; Indels 21; Gaps 5;

QY 32 LEGAPMGFTLKGLHGEPLIISKVEEGKADTLSSKLGADDEVHIN-EVTLSSRRE 90
 DB 7 LRGPMPGFRVLVGRDFSAPLTISRVAHAGSKA--ALALCPGDSIOAINGESTELTTLHE 64
 QY 91 AVSLVGSYKTLRLVVRSLSPPTVS-----LEPDQRQMRPRTTS--F 135
 DB 65 AQNRINGCHDLTLT---SVSRPNKWPSSPNDKAQRHRIHDPEADGSPATSRSSIS 121
 QY 136 SVSTADOR 143
 DB 122 GISLEDR 129

RESULT 8
 ID DLGI_RAT STANDARD; PRT; 911 AA.
 AC Q62656;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE PRESYNAPTIC PROTEIN SAP97 (SYNAPSE-ASSOCIATED PROTEIN 97) (DISCS,
 DE LARGE HOMOLOG 1).
 GN DLGI.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 ON NCBI_TaxID-10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Brain;
 RX MEDLINE-95198112; Pubmed-7891172;
 RA Mueller B.M., Kistner U., Voh R.W., Cases-Langhoff C., Becker B.,
 RA Gundelfinger E.D., Garner C.C.;
 RT "Molecular characterization and spatial distribution of SAP97, a
 RT novel presynaptic protein homologous to SAP90 and the Drosophila
 RT discs-large tumor suppressor protein."
 RL J. Neurosci. 15:2354-2366(1995).
 CC -1- FUNCTION: INTERACTS WITH THE CYTOPLASMIC TAIL OF NMDA RECEPTOR
 CC SUBUNITS. MAY BE INVOLVED IN SYNAPTOGENESIS. MAY PLAY A ROLE IN
 CC CELL ADHESION.
 CC -1- SUBCELLULAR LOCATION: IN CULTURED T84 CELLS, IT IS RESTRICTED TO
 CC THE CYTOPLASMIC SURFACE OF THE PLASMA MEMBRANES BETWEEN ADJACENT
 CC CELLS, BUT NOT AT THE EDGES OF CELLS LACKING CELL-CELL CONTACT.
 CC -1- TISSUE SPECIFICITY: PRESYNAPTIC NERVE TERMINI OF EXCITATORY
 CC SYNAPSES. IN OTHER BRAIN REGIONS, SAP97 IS FOUND IN AND ALONG
 CC BUNDLES OF UNMYELINATED AXONS. SAP97 IS NOT RESTRICTED TO THE CNS,
 CC BUT IS ALSO PRESENT AT THE BASAL LATERAL MEMBRANE BETWEEN A
 CC VARIETY OF EPITHELIAL CELLS.
 CC -1- SIMILARITY: CONTAINS 3 PDZ/DBP DOMAINS.
 CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
 CC -1- SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: U14950; AAA79976.1; -
 CC HSSP: Q12959; IPDR.
 CC InterPro: IPR000619; Guanylate_kin.
 CC InterPro: IPR001478; PDZ.
 CC InterPro: IPR001452; SH3.
 CC Pfam: PF00625; Guanylate_kin; 1.
 CC Pfam: PF00595; PDZ; 3.
 CC Pfam: PF00018; SH3; 1.
 CC SMART: SM00072; Gukc; 1.

```
DR SMART; SM00228; PDZ; 3.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS00856; GUANYLATE_KINASE_1; 1.
DR PROSITE; PS00502; GUANYLATE_KINASE_2; 1.
DR PROSITE; PS0106; PDZ; 3.
DR PROSITE; PS00326; SH3; 1.
DR PROSITE; PS00856; GUANYLATE_KINASE_1; 1.
DR PROSITE; PS00502; GUANYLATE_KINASE_2; 1.
KW SH3 domain; Repeat.
FT DOMAIN 224 310 PDZ 1.
FT DOMAIN 318 404 PDZ 2.
FT DOMAIN 465 545 PDZ 3.
FT DOMAIN 580 650 SH3.
FT DOMAIN 721 911 GUANYLATE_KINASE.
FT DOMAIN 527 530 POLY-ALA.
SQ SEQUENCE 911 AA; 100570 MW; 18CEBD31DD0CAF8B CRC64;

Query Match 11.8%; Score 121; DB 1; Length 911;
Best Local Similarity 25.8%; Pred. No. 0.0021;
Matches 42; Conservative 30; Mismatches 55; Indels 36; Gaps 7;

QY 3 RTEDFHKPSATLSNTATKGRYIYL-EAFL-----EGGAPWGFTLKGL 46
DB 424 QSVNDHVPSSVLTGTPASPARISPKAVLGDDDEITREPKVVLHKGSTGIGFNIVGG- 482
QY 47 ERGEPLIISKVEEGKADTLSSKQAGDEVVHINEVTLS-SRKEAVSLVKGSKYKTLRLVRSLSPPVTVSLE 118
DB 483 EDGEIGIFISFLLAGPAD-LSGELRKGDRISVNSVDLRAASHQAAALAKNA----- 534
QY 106 VRSLSPPVTVSLEFPDQHPQMPRP-----TRTSFVSVDADG 142
DB 535 ----GQAVTIAVQYRPEEYSFEAKIHDLRETMNSSVSSGG 573

RESULT 9
DLG3_MOUSE STANDARD; PRT; 849 AA.
AC P70175;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PRESYNAPTIC PROTEIN SAP102 (SYNAPSE-ASSOCIATED PROTEIN 102) (DISCS, LARGE HOMOLOG 3).
GN DLG3 OR DLG3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A. (LONG FORM).
RC TISSUE=Brain;
RA Kohnmura N., Makino S., Yagi T.;
RA Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.
RL FUNCTION: INTERACTS WITH THE CYTOPLASMIC TAIL OF THE NMDA RECEPTOR
CC SUBUNIT NR2B.
CC -1- SIMILARITY: CONTAINS 3 PDZ/DHR DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -1- SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D87117; BAAL3249.1; -
CC HSP; Q12959; LPDR.
CC MGD; MG1:188986; Dlg3.
CC InterPro; IPR000619; Guanylate_kin.
CC InterPro; IPR001478; PDZ.
CC InterPro; IPR001452; SH3.

DR Pfam; PF00625; Guanylate_kin; 1.
DR Pfam; PF00595; PDZ; 3.
DR Pfam; PF00018; SH3; 1.
DR SMART; SM00072; GUKC; 1.
DR SMART; SM00228; PDZ; 3.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS00856; GUANYLATE_KINASE_1; 1.
DR PROSITE; PS00502; GUANYLATE_KINASE_2; 1.
DR PROSITE; PS0106; PDZ; 3.
DR PROSITE; PS00002; SH3; 1.
KW SH3 domain; Repeat.
FT DOMAIN 149 235 PDZ 1.
FT DOMAIN 244 330 PDZ 2.
FT DOMAIN 404 484 PDZ 3.
FT DOMAIN 519 589 SH3.
FT DOMAIN 659 849 GUANYLATE_KINASE.
SQ SEQUENCE 849 AA; 93482 MW; EF3EF2D7513538EE CRC64;

Query Match 11.6%; Score 119.5; DB 1; Length 849;
Best Local Similarity 31.9%; Pred. No. 0.0027;
Matches 37; Conservative 24; Mismatches 44; Indels 11; Gaps 6;

QY 10 KPSATLSNTATKGRYIYLEAFLE-GGAPWGFTLKGLH-----GEP-LIISKVEEGSKA 63
DB 133 EPSLSVN---GSDGMFKYEEIVLGRNGLSGFAGGIDNPHVDPDPGIFITKIPGAA 189
QY 64 DTLSSKQAGDEVVHINEVTLS-SRKEAVSLVKGSKYKTLRLVRSLSPPVTVSLE 118
DB 190 -AMDGLGVNDCLVRNEVDVSEVHVSRAVEALKEAGPVRLVRRQPPPTIME 244

RESULT 10
DLG3_RAT STANDARD; PRT; 849 AA.
AC Q62936; P70547;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PRESYNAPTIC PROTEIN SAP102 (SYNAPSE-ASSOCIATED PROTEIN 102) (PSD-95/SAP90 RELATED PROTEIN 1) (DISCS, LARGE HOMOLOG 3).
GN DLG3.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A. (LONG FORM).
RC TISSUE=Brain;
RA MEDLINE=96374358; PubMed=8780649;
RA Mueller B.M., Kistner U., Kindler S., Chung W.J., Kuhlendahl S.,
RA Fenster S.D., Lau L.-F., Veh R.W., Haganir R.L., Gundelfinger E.D.,
RA Garner C.C.;
RA "SAP102, a novel postsynaptic protein that interacts with NMDA
RA receptor complexes in vivo.";
RA Neuron 17:255-265(1996).
RN [2]
RP SEQUENCE FROM N.A. (SHORT FORM).
RA Irie M., Hata Y., Takai Y.;
RA Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
RL FUNCTION: INTERACTS WITH THE CYTOPLASMIC TAIL OF THE NMDA RECEPTOR
CC SUBUNIT NR2B.
CC -1- SIMILARITY: CONTAINS 3 PDZ/DHR DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -1- SIMILARITY: CONTAINS A GUANYLATE_KINASE-LIKE DOMAIN.
CC -1- SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```

CC or send an email to license@lsb.slb.ch).
CC -----
CC EMBL: U50147; AAA93031.1;
CC DR EMBL: U53567; ABA48561.1;
CC DR HSSP: 012959; 1PDR.
CC DR InterPro: IPR000619; Guanylate_kin.
CC DR InterPro: IPR001478; PDZ.
CC DR InterPro: IPR001452; SH3.
CC DR Pfam: PF00625; Guanylate_kin. 1.
CC DR Pfam: PF00595; PDZ; 3.
CC DR Pfam: PF00018; SH3; 1.
CC DR SMART: SM00072; GUKC; 1.
CC DR SMART: SM00228; PDZ; 3.
CC DR SMART: SM00326; SH3; 1.
CC DR PROSITE: PS00856; GUANYLATE_KINASE_1; 1.
CC DR PROSITE: PS50052; GUANYLATE_KINASE_2; 1.
CC DR PROSITE: PS50106; PDZ; 3.
CC DR PROSITE: PS50002; SH3; 1.
CC DR SH3 domain; Repeat; Alternative splicing.
CC FT SH3 domain; 149 235 PDZ 1.
CC FT DOMAIN 244 330 PDZ 2.
CC FT DOMAIN 404 484 PDZ 3.
CC FT DOMAIN 519 589 SH3.
CC FT DOMAIN 659 849 GUANYLATE_KINASE.
CC FT VARSPLIC 627 640 MISSING (IN SHORT ISOFORM).
CC SEQUENCE 849 AA; 93539 MW; 34DA9C46C7BB96DB CRC64;

Query Match 11.6%; Score 119.5; DB 1; Length 849;
Best Local Similarity 31.9%; Pred. No. 0.0027;
Matches 37; Conservative 24; Mismatches 44; Indels 11; Gaps 6;

OY 10 KPSATINSNTATKRYVLEAFLE-GGAPMGFTLKGLEH---GEP-LIISKVEGGKA 63
Db 133 EPSLSVN---GSDGMFKVEELVLENGSGLGIDNPDPGIFIKIIPGGA 189
OY 64 DTLSSKLOAGDEVHINIVTSS-SRKAVALSVKSKYTLRLVYVSLPPTVSL 118
Db 190 -AMDRLGVNDCLRVNEVDYSEVHVSRAVLEKAGPYRLVVRROPPEITME 244

RESULT 11
ID PD1_HUMAN STANDARD; PRT; 329 AA.
AC 000151;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PDZ AND LIM DOMAIN PROTEIN 1 (LIM DOMAIN PROTEIN CLP-36) (C-TERMINAL
DE LIM DOMAIN PROTEIN 1) (ELFTN).
GN PD1IM1 OR CLIM1 OR CLP36.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OC NCBI_TaxID=9606;
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Heart;
RX MEDLINE=99145108; PubMed=10022510;
RA Lee C.Y., Ngai S.M., Garcia-Barcelo M., Tsui S.K.W., Fung K.P.,
RA Lee C.Y., Ngai S.M., Garcia-Barcelo M., Tsui S.K.W., Fung K.P.,
RA Ng E.K., Schaper J., Tsui S.K.W., Fung K.P., Lee C.Y., Ngai S.M.,
RA "Interaction of hclim1, an enigma family protein, with alpha-actinin
RT 2."
RT J. Cell. Biochem. 72:279-285(1999).
RN [2]
RP INTERACTION WITH ALPHA-ACTININ 2.
RX MEDLINE=20320865; PubMed=10861853;
RA Kotaka M., Koshin S., Ngai S., Chan K., Lau Y., Lee S.M., Li H.Y.,
RA Ng E.K., Schaper J., Tsui S.K.W., Fung K.P., Lee C.Y., Ngai S.M.,
RA "Interaction of hclim1, an enigma family protein, with alpha-actinin
RT 2."
RT J. Cell. Biochem. 78:558-565(2000).
RN [3]

```

```

RP INTERACTION WITH ALPHA-ACTININS 1 AND 4.
RX MEDLINE=20219155; PubMed=10753915;
RA Valienius T., Luukko K., Mäkelä T.P.;
RT CLP-36 PDZ-LIM protein associates with nonmuscle alpha-actinin-1 and
RT alpha-actinin-4.
RL J. Biol. Chem. 275:11100-11105(2000).
CC -1- FUNCTION: CYTOSKELETAL PROTEIN THAT MAY ACT AS AN ADAPTER THAT
CC BRINGS OTHER PROTEINS TO THE CYTOSKELETON.
CC -1- SUBUNIT: INTERACTS WITH ALPHA-ACTININS, 1 2 AND 4.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- TISSUE SPECIFICITY: STRONGLY EXPRESSED IN THE HEART AND SKELETAL
CC MUSCLE, MODERATELY EXPRESSED IN THE SPLEEN, SMALL INTESTINE,
CC COLON, PLACENTA, AND LUNG. A LOWER LEVEL EXPRESSION IS SEEN IN
CC LIVER, THYMUS, KIDNEY, PROSTATE AND PANCREAS AND IS NOT FOUND IN
CC THE BRAIN, TESTIS, OVARY, AND PERIPHERAL BLOOD LEUKOCYTES.
CC -1- SIMILARITY: CONTAINS 1 LIM DOMAIN. THE LIM DOMAIN BINDS 2 ZINC
CC IONS.
CC -1- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-slb.ch/announce/
CC or send an email to license@lsb.slb.ch).
CC -----
CC EMBL: U90878; AAC05580.1;
CC DR MIM: 605900;
CC DR InterPro: IPR001781; LIM.
CC DR InterPro: IPR001478; PDZ.
CC DR Pfam: PF00412; LIM; 1.
CC DR Pfam: PF00595; PDZ; 1.
CC DR Pfam: PF00094; LIM; 1.
CC DR Pfam: PF00094; LIM; 1.
CC DR SMART: SM00132; PDZ; 1.
CC DR SMART: SM00228; LIM; 1.
CC DR PROSITE: PS00478; LIM_DOMAIN_1; FALSE_NEG.
CC DR PROSITE: PS50023; LIM_DOMAIN_2; 1.
CC DR PROSITE: PS50106; PDZ; 1.
CC DR Cytoskeleton; LIM domain; Metal-binding; Zinc.
CC FT DOMAIN 3 84 PDZ.
CC FT DOMAIN 260 310 LIM.
CC SEQUENCE 329 AA; 36171 MW; 6520F1A932CFA312 CRC64;

Query Match 11.4%; Score 117.5; DB 1; Length 329;
Best Local Similarity 36.3%; Pred. No. 0.0013;
Matches 37; Conservative 14; Mismatches 40; Indels 11; Gaps 4;

OY 32 LEGAPMGFTLKGLEHEPLIISVKEEGKADTLSSKLOAGDEVHIN-EVTLSSSRKE 90
Db 8 LGGPGMGFTLKGLEHEPLIISVKEEGKADTLSSKLOAGDEVHIN-EVTLSSSRKE 65
OY 91 AVSLVGSYKTLRLVY-----RSLSPPTVSLIEPDQHPQRK 127
Db 66 AONRIKGTNDTLTVARESEHKVMSPLVY---EEGRHPYRK 104

RESULT 12
ID DLG1_HUMAN STANDARD; PRT; 904 AA.
AC 012959; 012958;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PRESYNAPTIC PROTEIN SAP97 (SYNAPSE-ASSOCIATED PROTEIN 97) (DISCS,
DE LARGE HOMOLOG 1).
GN DLG1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OC NCBI_TaxID=9606;
OX NCBI_TaxID=9606;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RX MEDLINE-95024052; PubMed-7937897;
RA Lue R.A., Marfatia S.M., Branton D., Chishti A.H.;
RT "Cloning and characterization of hdlg; the human homologue of the
RT Drosophila discs large tumor suppressor binds to protein 4.1.";
RN Proc. Natl. Acad. Sci. U.S.A. 91:9818-9822(1994).
RN [2]
RX X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS) OF 460-555.
RX MEDLINE-96338231; PubMed-87571139;
RA Cabral J.H., Petosa C., Sutcliffe M.J., Raza S., Byron O., Poy F.,
RA Marfatia S.M., Chishti A.H., Liddington R.C.;
RT "Crystal structure of a PDZ domain.";
RN Nature 382:649-652(1996).
CC -|- FUNCTION: INTERACTS WITH THE CYTOPLASMIC TAIL OF NMDA RECEPTOR
CC SUBUNITS. ASSOCIATES WITH PROTEIN 4.1.
CC -|- ALTERNATIVE PRODUCTS: 2 ISOFORMS; 1 (SHOWN HERE) AND 2; ARE
CC PRODUCED BY ALTERNATIVE SPLICING.
CC -|- SIMILARITY: CONTAINS 3 PDZ/DHR DOMAINS.
CC -|- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -|- SIMILARITY: CONTAINS A GUANYLATE KINASE-LIKE DOMAIN.
CC -|- SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; UI3897; AAA50599.1; -.
DR EMBL; UI3896; AAA50598.1; -.
DR PDB; 1PDR; 23-JUL-97.
DR MIM; 601014; -.
DR InterPro; IPR000619; Guanylate_kin.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00625; Guanylate_kin; 1.
DR Pfam; PF00595; PDZ; 3.
DR Pfam; PF00018; SH3; 1.
DR SMART; SM00072; GUKC; 1.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS00856; GUANYLATE_KINASE_1; 1.
DR PROSITE; PS0052; GUANYLATE_KINASE_2; 1.
DR PROSITE; PS50106; PDZ; 3.
DR PROSITE; PS50002; SH3; 1.
KW SH3 domain; Repeat; Alternative splicing; 3D-structure.
FT DOMAIN 224 310 PDZ 1.
FT DOMAIN 319 405 PDZ 2.
FT DOMAIN 466 546 PDZ 3.
FT DOMAIN 581 651 SH3.
FT DOMAIN 714 904 GUANYLATE KINASE.
FT VARSPLIC 669 680 EIPDGMGSKGLK -> QSFNDKRNKLNFSRKFFPKYKNQDS
FT EETSADQQ (IN ISOFORM 2).
FT SEQUENCE 904 AA; 100354 MW; B78798D6B80920D4 CRC64;

Query Match 11.0%; Score 113.5; DB 1; Length 904;
Best Local Similarity 26.2%; Pred. No. 0.01;
Matches 42; Conservative 27; Mismatches 54; Indels 37; Gaps 8;

QY 7 DFH-KPSATLNSNTATKGRYIYL-EAFL-----EGGAPMGFTLKGGLEHG 49
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 428 DNHVSPSFLGQTPASPARSPVSKAVLGDDDEITREPKVVLHRSSTGLGFNIYVG-EDG 486
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 50 EPLIISKVEGGKADTLSSKLOAGDEWHINEVTL-SSSRKEAVSLVKGSKYKTLRLVRS 108
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 487 EGIPFISILAGGPAD-LSGELRGKDRISVNSVDLRAASHEQAAAKNA----- 535
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 109 LSPPTVTSLEFDFQHPQMPMPRTR-----TSFSVSTADG 142
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 536 -GQATVIAQYREYRSRFEAKIHDLREQMNSSISGSG 574
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

```

RESULT 13
DLG3_HUMAN STANDARD; PRT; 817 AA.
ID DLG3_HUMAN STANDARD; PRT; 817 AA.
AC Q92796; OSUL18;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PRESYNAPTIC PROTEIN SAP102 (SYNAPSE-ASSOCIATED PROTEIN 102)
DE (NEUROENDOCRINE-DLG) (NE-DLG) (DISCS, LARGE HOMOLOG 3).
GN DLG3 OR KIAA1232
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=Fetal brain;
RX MEDLINE-97332623; PubMed-9188857;
RA Makino K., Kuwahara H., Masuko N., Nishiyama Y., Morisaki T.,
RA Sasaki J., Nakao M., Kuwano A., Nakata M., Ushio Y., Saya H.;
RT "Cloning and characterization of NE-dlg; a novel human homologue of the
RT Drosophila discs large (dlg) tumor suppressor protein interacts with
RT the APC protein.";
RL Oncogene 14:2425-2433(1997).
RN [2]
RP SEQUENCE OF 330-817 FROM N.A.
RX TISSUE=Brain;
RX MEDLINE-20039619; PubMed-10574462;
RA Nagase T., Ishikawa K.-I., Kikuno R., Hirose M., Nomura N.,
RA Chara O.;
RT "Prediction of the coding sequences of unidentified human genes. XV.
RT The complete sequences of 100 new cDNA clones from brain which code
RL DNA Res. 6:337-345(1999).
CC -|- FUNCTION: INTERACTS WITH THE CYTOPLASMIC TAIL OF THE NMDA RECEPTOR
CC SUBUNIT NR2B (BY SIMILARITY).
CC -|- SIMILARITY: CONTAINS 3 PDZ/DHR DOMAINS.
CC -|- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -|- SIMILARITY: CONTAINS A GUANYLATE KINASE-LIKE DOMAIN.
CC -|- SIMILARITY: BELONGS TO THE MAGUK FAMILY OF CELL JUNCTION PROTEINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U49089; AAB61453.1; -.
DR EMBL; AB033058; BAA86546.1; -.
DR HSSP; O12959; 1PDR.
DR MIM; 300189; -.
DR InterPro; IPR000619; Guanylate_kin.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00625; Guanylate_kin; 1.
DR Pfam; PF00595; PDZ; 3.
DR Pfam; PF00018; SH3; 1.
DR SMART; SM00072; GUKC; 1.
DR SMART; SM00228; PDZ; 3.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS00856; GUANYLATE_KINASE_1; 1.
DR PROSITE; PS50106; PDZ; 3.
DR PROSITE; PS50002; SH3; 1.
KW SH3 domain; Repeat.
FT DOMAIN 130 217 PDZ 1.
FT DOMAIN 226 311 PDZ 2.
FT DOMAIN 379 465 PDZ 3.
FT DOMAIN 503 568 SH3.

```

FT DOMAIN 628 803 GUANYLATE KINASE.
 FT CONFLICT 330 381 FTALADNHSHNSISGLGAVESKVSYPAPPOVPTRYSPY
 FT CONFLICT 330 381 PRHIAEDFT -> AARRGAMERARKFSGLIAGLS
 FT CONFLICT 592 606 ASASARRASORAMPLRLSLPGDA (IN REF. 2).
 FT CONFLICT 592 606 DFGLSDDYGAHKL -> SIKTRKRSFSLRFPYKSK
 FT SEQUENCE 817 AA; 90344 MW; 3D7512EC4713FC4E CRC64;
 ENMAQESSIDOGVTSNTSDESS (IN REF. 2).

Query Match 10.9%; Score 112; DB 1; Length 817;
 Best Local Similarity 33.0%; Pred. No. 0.013;
 Matches 34; Conservative 20; Mismatches 41; Indels 8; Gaps 5;

QY 23 GRVYLFLEAFLE-GGAPMGFTLKGLEH---GEP-LIISKVEEGKADTLSSKLGAGDEV 76
 DB 125 GMFEYEEVLEIRNGSGFSLAGIDNPVDDPGIFITIKIIPGAA-AMDGRLGVNDCV 183
 QY 77 VHNVEVTLSS-SRKEAVSLVGSYKTLRLVYRSLSPVTVLE 118
 DB 184 LRNVEVEVEVHSRAVEALKAGPVRLVYRRQPPETIME 226

RESULT 14

RIL_HUMAN STANDARD; PRT; 328 AA.
 ID RIL_HUMAN
 AC P50479;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE LIM PROTEIN RIL (REVERSION-INDUCED LIM PROTEIN).
 GN RIL.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Lung;
 RA Schart B., Schaefer R.;
 RL Submitted (OCT-1995) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE OF 1-20 FROM N.A.
 RX MEDLINE=98248906; Pubmed=9573374;
 RA Bashitova A.A., Markelov M.L., Shlykova T.V., Levshenkova E.V.,
 RA Alibaeva R.A., Frolova E.I.;
 RT "The human RIL gene: mapping to human chromosome 5q31.1, genomic
 RT organization and alternative transcripts.";
 RL Gene 210:239-245(1998).

-1- SIMILARITY: CONTAINS 1 LIM DOMAIN. THE LIM DOMAIN BINDS 2 ZINC
 IONS.
 CC -1- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).

CC EMBL: X93510; CAA63767.1; -
 DR EMBL: U82997; AAC52072.1; -
 DR HSSP: Q12959; 1PDR.
 DR MIM: 603422; -
 DR InterPro: IPR001781; LIM.
 DR InterPro: IPR001478; PDZ.
 DR Pfam: PF00412; LIM; 1.
 DR Pfam: PF00595; PDZ; 1.
 DR ProDom: PD000094; LIM; 1.
 DR ProDom: SM00132; LIM; 1.
 DR SMART: SM00132; LIM; 1.
 DR SMART: SM00228; PDZ; 1.
 DR PROSITE: PS00478; LIM_DOMAIN_1; 1.
 DR PROSITE: PS50023; LIM_DOMAIN_2; 1.
 DR PROSITE: PS50023; LIM_DOMAIN_2; 1.

DR PROSITE: PS50106; PDZ; 1.
 KW LIM domain; Metal-binding; zinc.
 FT DOMAIN 8 83 PDZ.
 FT DOMAIN 253 303 LIM.
 SO SEQUENCE 328 AA; 35082 MW; B0AF866B1879E747 CRC64;

Query Match 10.7%; Score 110; DB 1; Length 328;
 Best Local Similarity 28.0%; Pred. No. 0.0062;
 Matches 46; Conservative 18; Mismatches 60; Indels 40; Gaps 8;

QY 32 LEGAPMGFTLKGLEHGEPLISKVEEGKADTLSSKLGAGDEVVHIN-EVTLSSRKE 90
 DB 7 LRGPSPWGFRLV-GRDSAPLTIKRVHAGSKASL--AALCEDLIQAINGESTELMTHLE 63
 QY 91 AVSLVGSYKTLRLVYRSLSPVTVS-----LEFPDHPORMPPTRTSFS 136
 DB 64 AONRIKGDHLLT---SVSRPEGRSWPSAPDDSKAQRHIDPEIDGSPITSRRPSG 120
 QY 137 VSTA--DGRHEMSCRPVYKWSRPPTWAKRPQKCIYPOHN 178
 DB 121 TGTGPEDEG-----RPSLGSPYGRKPC-FPVPHN 147

RESULT 15

PDL1_MOUSE STANDARD; PRT; 326 AA.
 ID PDL1_MOUSE
 AC O70400;
 DT 20-AUG-2001 (Rel. 40, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE PDZ AND LIM DOMAIN PROTEIN 1 (LIM DOMAIN PROTEIN CLP-36) (C-TERMINAL
 DE LIM DOMAIN PROTEIN 1) (ELF1N).
 GN PDLIM1 OR CLIM1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_Taxid=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Kotaka M., Ngai S.M., Tsui S.K.W., Fung K.P., Lee C.Y., Waye M.M.Y.;
 RT "Characterization of mouse 36kDa carboxyl terminal LIM domain protein
 RT (mCLIM1).";

RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: CYTOSKELETAL PROTEIN THAT MAY ACT AS AN ADAPTER THAT
 CC BRINGS OTHER PROTEINS TO THE CYTOSKELETON (BY SIMILARITY).
 CC -1- SUBUNIT: INTERACTS WITH ALPHA-ACTININS, 1 2 AND 4 (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
 CC -1- SIMILARITY: CONTAINS 1 LIM DOMAIN. THE LIM DOMAIN BINDS 2 ZINC
 IONS.
 CC -1- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).

CC EMBL: AF053367; AAC08436.1; -
 DR EMBL: MG1:1860611; CLIM1.
 DR InterPro: IPR001781; LIM.
 DR InterPro: IPR001478; PDZ.
 DR Pfam: PF00412; LIM; 1.
 DR Pfam: PF00595; PDZ; 1.
 DR ProDom: PD000094; LIM; 1.
 DR ProDom: SM00132; LIM; 1.
 DR SMART: SM00228; PDZ; 1.
 DR SMART: SM00478; LIM_DOMAIN_1; FALSE_NEG.
 DR PROSITE: PS50023; LIM_DOMAIN_2; 1.
 DR PROSITE: PS50023; LIM_DOMAIN_2; 1.
 DR PROSITE: PS50106; PDZ; 1.
 KW Cytoskeleton; LIM domain; Metal-binding; zinc.

FT DOMAIN 3 84 PDZ.
FT DOMAIN 257 307 LIM.
SQ SEQUENCE 326 AA; 35717 MW; 5228220E59C1A90E CRC64;

Query Match 10.6%; Score 109; DB 1; Length 326;
Best Local Similarity 36.3%; Pred. NO. 0.0076;
Matches 37; Conservative 14; Mismatches 39; Indels 12; Gaps 5;

QY 32 LEGGAPWFTLKGLEHGEPLIISKVEGGKADTLSSKLOAGDEVVHIN-EVTLSSSRKE 90
|:|:||||| | | : :||| ||| | | | : | | | : | | | | | | |

Db 8 LOGPGWGFRLV-GKDFEQPLAISRVTPGSKA--AIANLCIGLITAIIDGEDTSSMTHLE 64
| : | | | : | | | : | | | | | | | | | | | | |

QY 91 AVSLVRGSKYKTLRLVY-----RSLSPPTVVSLEFDPQHPQRM 127
| : | | | : | | | : | | | | | | | | | | | | |

Db 65 AQNKIKGCADNMTLTVSRSEQKINSPLVT---EGRKRHPYKM 103
| : | | | : | | | : | | | | | | | | | | | | |

Search completed: January 31, 2002, 21:06:28
Job time: 51 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2002, 21:03:57 ; Search time 72.88 seconds
(without alignments)
381.336 Million cell updates/sec

Title: US-09-641-831-2

Perfect score: 1029
Sequence: 1 MRRTEDEHFKPSATLNSNTA.....CIYPQHNTGCRKRAVLSR 190

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: SP-archaea:*
2: SP-bacteria:*
3: SP-fungi:*
4: SP-human:*
5: SP-invertebrate:*
6: SP-mammal:*
7: SP-mhc:*
8: SP-organelle:*
9: SP-phage:*
10: SP-plant:*
11: SP-rodent:*
12: SP-virus:*
13: SP-vertebrate:*
14: SP-unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	426	41.4	1986	11	Q9GXNO mus musculu
2	271.5	26.4	1502	4	Q9DIL8 Q9H4G2
3	143.5	13.9	1454	4	Q9H4G2
4	138.5	13.5	298	13	Q9PU46 Q9Y423
5	138.5	13.5	315	13	Q9PU47
6	136	13.2	470	4	Q9Y423
7	135.5	13.2	661	11	Q9Y423
8	135.5	13.2	723	11	Q9Y423
9	135.5	13.2	723	11	Q9Y423
10	133.5	13.0	171	5	Q17458
11	132.5	12.9	234	11	Q9D2Q1
12	131	12.7	283	4	Q9Y423
13	131	12.7	288	11	Q9Y423
14	131	12.7	288	11	Q9Y423
15	131	12.7	536	4	Q9Y423
16	131	12.7	617	4	Q9Y423
17	131	12.7	1278	4	Q9Y423
18	131	12.7	1455	4	Q9Y423
19	130.5	12.7	138	11	Q9D130

20	130.5	12.7	493	5	Q9W443	Q9W443 drosophila
21	130.5	12.7	734	4	Q9Y112	Q9Y112 homo sapien
22	129.5	12.6	207	11	Q9Z252	Q9Z252 rattus norv
23	128.5	12.5	207	4	Q9HAP6	Q9HAP6 homo sapien
24	128.5	12.5	400	5	Q9V7F2	Q9V7F2 drosophila
25	128	12.4	39	4	Q9H4T0	Q9H4T0 homo sapien
26	128	12.4	435	4	Q14250	Q14250 homo sapien
27	127.5	12.4	316	11	Q70209	Q70209 mus musculu
28	127	12.3	591	11	Q9QYN0	Q9QYN0 mus musculu
29	127	12.3	337	11	Q9QYN2	Q9QYN2 mus musculu
30	126.5	12.3	207	4	Q9H4T0	Q9H4T0 homo sapien
31	126.5	12.3	403	4	Q9H4T0	Q9H4T0 homo sapien
32	126.5	12.3	552	4	Q9H4T0	Q9H4T0 homo sapien
33	126.5	12.3	652	4	Q9H4T0	Q9H4T0 homo sapien
34	126	12.2	195	5	Q9VBY7	Q9VBY7 drosophila
35	125.5	12.2	327	11	Q9KX93	Q9KX93 mus musculu
36	125.5	12.2	329	4	Q9BXP2	Q9BXP2 homo sapien
37	124.5	12.1	153	4	Q9BXP2	Q9BXP2 homo sapien
38	124.5	12.1	214	11	Q9CRA2	Q9CRA2 mus musculu
39	124.5	12.1	457	4	Q9H4T0	Q9H4T0 homo sapien
40	124.5	12.1	521	4	Q9H4T0	Q9H4T0 homo sapien
41	124	12.1	591	11	Q62920	Q62920 rattus norv
42	123.5	12.0	548	11	Q9ES65	Q9ES65 mus musculu
43	123.5	12.0	910	11	Q9ES64	Q9ES64 mus musculu
44	122.5	11.9	189	4	Q9BVP4	Q9BVP4 homo sapien
45	122	11.9	362	11	Q70208	Q70208 rattus norv

ALIGNMENTS

RESULT 1
ID Q9GXNO PRELIMINARY: PRT: 1986 AA.
AC Q9GXNO;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE PDZ DOMAIN ACTIN BINDING PROTEIN SHROOM.
GN SHRM.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathu; Muridae; Murinae; Mus.
CX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A. PubMed-10589677;
RX MEDLINE-2005594; Soriano P.;
RA "Shroom, a PDZ domain-containing actin-binding protein, is required
RT for neural tube morphogenesis in mice."
RL Cell 99:485-497(1999).
DR EMBL: AF199421; AAF13269.1; -
DR HSRP: P29476; ICAV.
DR MGI: 1351655; Shrm.
DR InterPro: IPR001478; PDZ.
DR Pfam: PF00595; PDZ_1.
DR SMART: SM00228; PDZ_1.
DR PROSITE: PS50106; PDZ_1.
SQ SEQUENCE 1986 AA; 215261 MW; 18C5EFCB772EABF CRC64;

Query Match 41.4%; Score 426; DB 11; Length 1986;
Best Local Similarity 46.3%; Pred. No. 6.3e-30;
Matches 101; Conservative 24; Mismatches 57; Indels 36; Gaps 6;
QY 2 MRRTEDEHFKPSATLNSNTA...KRYIYLFEGAPWGTLLKGLGHEGPEPLIKVEEG 61
DB 1 MRRTEDEHFKPSATLNSNTA...KRYIYLFEGAPWGTLLKGLGHEGPEPLIKVEEG 60
QY 62 KADTSKQAGDEVEHINEVTLSSRKRAVSLVKGSKYTLRLVVR-----SLSPV 113
DB 61 KADTSKQAGDEVEHINEVTLSSRKRAVSLVKGSKYTLRLVVRDVCAPGHADPGT 120

```
QY 114 TVSLFED-----POHPQ-----RMPRTRTSFVSADGRHEWSCRPVWVWWSRPRT 161
DB 121 SKSLSELLTSPQRKATGCGVKLRKQKCS---EPATRPHSWHT-----TKGEIQPD 173
QY 162 WAARWPQKQCIYPTQHT-----CRNFKRAYLSR 190
DB 174 VSMQISQGTGMPWPHQSHSSSTSDLSNDYHAYLR 211

RESULT 2
ID Q9ULL8 PRELIMINARY; PRT; 1502 AA.
AC Q9ULL8;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DE 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE KIAA1202 PROTEIN (FRAGMENT).
GN KIAA1202.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=Brain;
RX MEDLINE=20039619; PubMed=10574462;
RA Nagase T., Ishikawa K., Kikuno R., Hirose M., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XV.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 6:337-345(1999).
DR EMBL; AB033028; BAA86516.1; -.
DR HSSP; P31016; 1BE9.
DR InterPro; IPR001478; PDZ.
DR Pfam; PF00595; PDZ; 1.
DR SMART; SM00228; PDZ; 1.
DR PROSITE; PS50106; PDZ; 1.
FT NON_TER 1
SQ SEQUENCE 1502 AA; 165950 MW; DC050C9C69B78AD9 CRC64;

Query Match 26.4%; Score 271.5; DB 4; Length 1502;
Best Local Similarity 56.5%; Pred. No. 5.4e-16;
Matches 52; Conservative 19; Mismatches 20; Indels 1; Gaps 1;

QY 23 GRYILEAFLEGAPWGFTLKGLEHGEPLIISKVEEGKADTLSSKLOAGDEVVHINEV 82
DB 10 GSFQYVPVQLGGAPWGFTLKGLEHGEPLIYSKIEDGGKA-ALSKMRTGDELVNINGT 68
QY 83 TLSSSRKEAVSLVKGSKYKTLRLVVRSLSPVT 114
DB 69 PLYGSRQREALILIKGSRILKLVRRNAPVS 100

RESULT 3
QYH4G2 PRELIMINARY; PRT; 1454 AA.
ID QYH4G2;
AC QYH4G2;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE BAI19E20.1 (KIAA1202 PROTEIN) (FRAGMENT).
GN BAI19E20.1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Wilson S.;
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL121865; CAC08472.1; -.
DR InterPro; IPR001478; PDZ.
```

```
DR SMART; SM00228; PDZ; 1.
DR PROSITE; PS50106; PDZ; 1.
FT NON_TER 1
SQ SEQUENCE 1454 AA; 160615 MW; 7931571D7DA80DE8 CRC64;

Query Match 13.9%; Score 143.5; DB 4; Length 1454;
Best Local Similarity 50.0%; Pred. No. 0.00024;
Matches 29; Conservative 15; Mismatches 13; Indels 1; Gaps 1;

QY 57 VEEGKADTLSSKLOAGDEVVHINEVTLSSSRKEAVSLVKGSKYKTLRLVVRSLSPVT 114
DB 1 IEDGGKA-ALSKMRTGDELVNINGTPLYGSRQREALILIKGSRILKLVRRNAPVS 57

RESULT 4
QYPU46 PRELIMINARY; PRT; 298 AA.
ID QYPU46;
AC QYPU46;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE ALPHA-ACTININ ASSOCIATED LIM PROTEIN, SKELETAL MUSCLE ISOFORM (SKALP)
DE (FRAGMENT).
GN ALP.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99436131; PubMed=10506181;
RA Ponies P., Macalma T., Beckerle M.C.;
RT "Purification and characterization of an alpha-actinin-binding PDZ-LIM
RT protein that is up-regulated during muscle differentiation.";
RL J. Biol. Chem. 274:29242-29250(1999).
DR EMBL; AJ249219; CAB53971.1; -.
DR InterPro; IPR001478; PDZ.
DR Pfam; PF00595; PDZ; 1.
DR PROSITE; PS50106; PDZ; 1.
DR SMART; SM00228; PDZ; 1.
FT NON_TER 298
SQ SEQUENCE 298 AA; 32094 MW; 27D6F24FC6141E61 CRC64;

Query Match 13.5%; Score 138.5; DB 13; Length 298;
Best Local Similarity 31.7%; Pred. No. 9.2e-05;
Matches 38; Conservative 21; Mismatches 38; Indels 23; Gaps 5;

QY 32 LEGGAPWGFTLKGLEHGEPLIISKVEEGKADTLSSKLOAGDEVVHINEV-TLSSSRKE 90
DB 7 LFGAPWGFRSLGGIDFNQPLIITRITPGSKAST--ANLCPGDIIVAINGLSTENHTND 64
QY 91 AVSLVKGSKYKTLRLVVRSL-----SP-----PVTVSLEFPDQ-----HPQRMPPR 130
DB 65 AGERIAAAHQSLRKRAETKLNSQVSDGKANKYKINLEAFQDINFEHRHNIRPK 124

RESULT 5
QYPU47 PRELIMINARY; PRT; 315 AA.
ID QYPU47;
AC QYPU47;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE ALPHA-ACTININ ASSOCIATED LIM PROTEIN, SMOOTH MUSCLE ISOFORM (SMALP).
GN ALP.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
```


SQ SEQUENCE 315 AA; 34365 MW; 86414680A89BCA35 CRC64;
 DR LIM motif, metal-binding; Zinc.
 DR PROSITE: PS50106; PDZ: 1.
 DR SMART: SM00228; PDZ: 1.
 DR PROSITE: PS50023; LIM_DOMAIN_2; 1.
 DR SMART: SM00132; LIM: 1.
 DR PRODOM: PD000094; LIM: 1.
 DR Pfam; PF00585; PDZ: 1.
 DR Pfam; PF00412; LIM: 1.
 DR InterPro: IPR001478; PDZ.
 DR InterPro: IPR001781; LIM.
 DR EMBL: AJ249218; CAB53970.1; -.
 CC IONS.
 CC -1- SIMILARITY: CONTAINS LIM DOMAIN(S). THE LIM DOMAIN BINDS 2 ZINC
 RT protein that is up-regulated during muscle differentiation.";
 RT J. Biol. Chem. 274:29242-29250(1999).
 RA Pontier P., MacIama T., Beckerle M.C.;
 RX MEDLINE-99436131; Pubmed-10506181;
 [1] SEQUENCE FROM N.A.

[illegible]

```

RESULT      6
ID          09Y423
AC          09Y423
DT          01-NOV-1999 (TREMBLrel. 12, Created)
DT          01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT          01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE          ZASP PROTEIN (FRAGMENT).
GN          ZASP.
OS          Homo sapiens (Human).
OC          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC          Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX          NCBI_Taxid=9606;
RN          [1]
RP          SEQUENCE FROM N.A.
RC          TISSUE-SKELETAL MUSCLE, AND HEART;
RA          Faulner G., Pallavicini A., Formetlin E., Comelli A., Ievolella C.,
RA          Trevisan S., Bortoletto G., Scannapieco P., Salamon M., Mouly G.,
RA          Valle G., Lanfranchi G.;
RT          *ZASP: a new z-band alternatively spliced PDZ-motif protein.*
RL          J. Cell Biol. 0:0-0(0).
CC          -1 SIMILARITY: CONTAINS LIM DOMAIN(S). THE LIM DOMAIN BINDS 2 ZINC
CC          IONS.
DR          EMBL: AJ133768; CAB46729.1; -.
DR          HSSP: Q05158; IQLT.
DR          InterPro: IPR003006; Iq_MHC.
DR          InterPro: IPR001781; LIM.
DR          InterPro: IPR001478; PDZ.
DR          InterPro: IPR002965; P_Rich_extensn.
DR          Pfam: PF00412; LIM; 3.
DR          Pfam: PF00595; PDZ; 1.
DR          PRINTS: PR01217; PRICHEXTENS.
DR          ProDom: PD000094; LIM; 3.

```

DR SMART; SMO0312; LIM; 3
DR SMART; SMO0228; PDZ; 1
DR PROSITE; PS00290; TG_MHC; UNKNOWN_1
DR PROSITE; PS00478; LIM_DOMAIN_1; 2
DR PROSITE; PS50023; LIM_DOMAIN_2; 3
DR PROSITE; PS5106; PDZ; 1
DR LIM motif; Metal-binding; zinc
FM NON_TER 1
SQ SEQUENCE 470 AA; 50661 MW; AEF87648FA1A2BCF CRC64;

[illegible]

Q9JKS3	PRELIMINARY;	PRT;	661 AA.
ID	Q9JKS3		
AC	Q9JKS3		
DT	01-OCT-2000 (TrEMBLrel. 15, Created)		
DI	01-OCT-2000 (TrEMBLrel. 15, Last sequence update)		
DR	01-JUN-2001 (TrEMBLrel. 17, last annotation update)		
DE	ORACLE 2 PROTEIN (FRAGMENT).		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBI_Taxid:10090;		
PN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=20193503; PubMed=10727866;		
RA	Paslier R., Richardson J.A., Olson E.N.;		
RT	"Oracle, a novel PDZ-LIM domain protein expressed in heart and		
RL	skeletal muscle."		
RT	Mech. Dev. 92:277-284 (2000).		
CC	-1 SIMILARITY: CONTAINS LIM DOMAIN(S). THE LIM DOMAIN BINDS 2 ZINC		
CC	IONS.		
DR	EMBL; AF228058; AAF3848.1; "		
DR	InterPro: IPR000345; CYTC_Heme_bind.		
DR	InterPro: IPR003006; 1g_MTC.		
DR	InterPro: IPR001781; LIM.		
DR	InterPro: IPR001478; PDZ.		
DR	InterPro: IPR002965; P_rich_extensn.		
DR	Pfam; PF00412; LIM; 3.		
DR	Pfam; PF00593; PDZ; 1.		
DR	PRINTS; PRO1217; PRICEXTENS.		
DR	ProDom; PDO00094; LIM; 3.		
DR	SMART; SM00132; LIM; 3.		
DR	SMART; SM00228; PDZ; 1.		
DR	PROSITE; PS00190; CYTOCHROME_C; UNKNOWN_1.		
DR	PROSITE; PS00290; IG_MHC; UNKNOWN_1.		
DR	PROSITE; PS00478; LIM_DOMAIN_1; 2.		
DR	PROSITE; PS50023; LIM_DOMAIN_2; 3.		
DR	PROSITE; PS50106; PDZ; 1.		
KW	LIM motif; Metal-binding; zinc.		
FT	NON_TER 661		
SQ	SEQUENCE 661 AA. 70750 MW. 5DFEEFE78DA5532F CRC64;		

Query Match	13.2%;	Score 135.5;	DB 11;	Length 661;
Best Local Similarity	31.7%;	Pred. No. 0.00048;		

RESULT	9	
Q9VWH2		
ID	Q9VWH2	PRELIMINARY;
AC	Q9VWH2;	PRT;
DT	01-NOV-1999	723 AA.
DT	01-MAR-2001	
DT	01-NAR-2001	(TRENBLrel. 12, Created)
DT	01-JUN-2001	(TRENBLrel. 16, Last sequence update)
DE	CYPHER1.	
DE	CYPHER1.	(TRENBLrel. 17, Last annotation update)
DE	ZASP.	
OS	MOUSE	
OS	MOUSE	

OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CC	NCBI_TaxID=10090;
RN	[1]
RN	SEQUENCE FROM N.A.
RP	MEDLINE=99321910; PubMed=10391924;
RX	Zhou Q., Ruiz-Lozano P., Martone M.E., Chen J.;
RT	"Cypher, a Striated Muscle-restricted PDZ and LIM Domain-containing
RT	Protein, Binds to alpha-Actinin-2 and Protein Kinase C.";
RT	J. Biol. Chem. 274:19807-19813(1999).
RL	[2]
RL	SEQUENCE FROM N.A.
RP	Chen J., Zhou Q.;
RA	Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
RL	-!- SIMILARITY: CONTAINS LIM DOMAIN(S). THE LIM DOMAIN BINDS 2 ZINC
CC	IONS.

Ddb		65 AQNKIKSASYNLSLTLOKSKRPISTAPQSPLEVPVPHOKDPALDINGSLATPSPSP 12
OY		145 EWSGRP 150
Db		125 EARASP 130
RESULT 10		
ID	017458	PRELIMINARY; PRT; 171 AA.
IC	AD	
AC	017458;	
DT	01-JAN-1998	(TREMBLrel. 05, Created)
DT	01-JAN-1998	(TREMBurel. 05, Last sequence update)
DT	01-JUN-2001	(TREMBrel. 17, Last annotation update)

	Query Match	13.2%	Score 135.5	DB 11	Length 723
	Best Local Similarity	31.7%	Pred. No. 0.00053		
	Matches	40	Conservative	19	Mismatches 58
					Indels 9
					Gaps
QY	32	LEGGAPWGFTLKGLEHGEPLIIKSGVEEGKADTLGSKLQAGDEVVHINEV-TLSSRRKE	90		
Db	7	LTGPGWCFRGLGGKDFENMPLTISRTPGSKA--AQSLSQGLVVAIDGVNTDWTMLE	64		
QY	91	AVSLVKGSYKYLRLVWLSLSPPTVTSLEFPDQH-----PQMPPTRTTSFVSVDAGRH	14		
Db	65	AQNKIKSAYNSLTQAKSRKRPPISTAPTQISPLVPFHOKDPALDTNGSLVPSFSP	12		

QY	32	LEGGAPWGFTLKGGLHEGEPLIISKVEEGKADTLSSKLQAGDEVVHNEV-TLSSSRKE	90
		<pre> + + + + + : : : : : : : : : : : : : : : + + + + + : : : : : : : : : : : : : : + + + + + : : : : : : : : : : : : : : </pre>	
Db	7	LTGPGPWGRLOQGKDFNPLTISRTPGSKA--AQSQLSGDLVVAIDGVNFDMTHTLE	64

QY	91	AVSLVKGSYKTLRLVWRSLSPPVTVSLEFPQH-----PQRMPRTTSFVSVDAGRH 14
Ddb	65	AQNLIKASYNLSLTLOKSRPIPISTAPPISPLFIHQKDPALDNGSLATPSFP 12
QY	145	EWSCRP 150
Dd	125	EARASP 130

RESULT	10
ID	O17458
PRT;	PRELIMINARY; PRT; 171 AA.
AC	O17458; (TREMBlrel. 05, Created)
DT	01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DDT	01-JUN-1998 (TREMBlrel. 05, Last annotation update)
DCT	01-JUN-2001 (TREMBlrel. 17, Last annotation update)

DE HYPOTHETICAL 19.0 KDA PROTEIN (FRAGMENT).
 OS Schistosoma mansoni (Blood fluke).
 OC Eukaryota; Metazoa; Platyhelminthes; Rhabdophora; Neodermata;
 OC Trematoda; Digenea; Strigoida; Schistosomatidae; Schistosomatidae;
 OC Schistosoma.
 NCBI_TaxID=6183;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=PUERTO RICAN;
 RA Harrop R., Wilson R.A.;
 RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF030965; AAB86566.1; -
 DR HSSP; P31016; 1BFE.
 DR InterPro; IPR001478; PDZ.
 DR Pfam; PF00595; PDZ; 1.
 DR SMART; SM00228; PDZ; 1.
 DR PROSITE; PS50106; PDZ; 1.
 KW Hypothetical protein.
 FT NON_TER 1
 SQ SEQUENCE 171 AA; 19000 MW; 69685F4108598063 CRC64;

Query Match 13.0%; Score 133.5; DB 5; Length 171;
 Best Local Similarity 33.8%; Pred. No. 0.00013;
 Matches 47; Conservative 24; Mismatches 47; Indels 21; Gaps 8;

QY 4 TTEDFH-----KPSATLSN-----TATKG---RYIYLEAFLEGAPWGTLLKGLGEHGE 50
 Db 34 TTVYDNGSEEVKASATATVAFAFASGHAPRYELPKTNEG---LGFNVMGKEQNS 90
 QY 51 PLISYVEEGKADTLLSKLQAGDEVVHINEVTLSSSRKE-AVSLVKGSKYKTLRVVSL 109
 Db 91 PIYISRMKPGGVADR-HGGLKRGDQLSYNGISVESEHHEKAVELLKLAQGVKLVR-Y 148
 QY 110 SPPTVSL--FDPQHPQR 126
 Db 149 TPRIIEEAPDKOKARR 167

RESULT 11
 Q9D201 PRELIMINARY; PRT; 234 AA.

AC Q9D201 01-JUN-2001 (TREMBLrel. 17, Created)
 DC 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE VERTEBRATE HOMOLOG OF C. ELEGANS LIN-7 TYPE 2.
 GN VEL12.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=HIPPOCAMPUS;
 RL MEDLINE-21083660; PubMed-11217851;
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 Aikawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamahata I.,
 Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 Kadota K., Matsuda H.A., Ashburner M., Batilov S., Casavant T.,
 Fletschmann W., Gaasterland T., Glassl C., King B., Kochiya H.,
 Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
 Schiml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 Sasakhi H., Sato K., Schenbach C., Seya T., Shibata Y., Storch K.-F.,
 Suzuki H., Taya-Oka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
 Wyszynski-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,

RA Hayashizaki Y.;
 RL "Functional annotation of a full-length mouse cDNA collection.";
 RL Nature 409:685-690(2001).
 DR EMBL; AK019299; BAB31655.1; -
 DR MGD; MG1:1330858; Vel12.
 DR InterPro; IPR001478; PDZ.
 DR Pfam; PF00595; PDZ; 1.
 DR SMART; SM00228; PDZ; 1.
 DR PROSITE; PS50106; PDZ; 1.
 SQ SEQUENCE 234 AA; 25732 MW; 76919D8B1C29E53E CRC64;

Query Match 12.9%; Score 132.5; DB 11; Length 234;
 Best Local Similarity 32.9%; Pred. No. 0.00024;
 Matches 47; Conservative 25; Mismatches 44; Indels 27; Gaps 8;

QY 13 ATLNSNTATKG---RYIYLEAFLEGAPWGTLLKGLGEHGEPLISYVEEGKADTLLS 68
 Db 77 ATVAFAFASGHAPRYVELPKTNEG---LGFNVMGKEQNSPIYISHVIFGVADR-HG 132
 QY 69 KLAGDEVVHINEVTLSSSRKE-AVSLVKGSKYKTLRVVSLSPPTVSL--FD----- 120
 Db 133 GLKRGDQLSYNGISVESEHHEKAVELLKLAQGVKLVR-YTPRVLEMEAREKRSRA 191
 QY 121 ---PQHPQ-----RMPDPTRT 133
 Db 192 RRRPAPQLHVLGVSRKLSQTRT 214

RESULT 12
 Q9Y4Z5 PRELIMINARY; PRT; 283 AA.

AC Q9Y4Z5 01-NOV-1999 (TREMBLrel. 12, Created)
 DC 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE ZASP PROTEIN (FRAGMENT).
 GN ZASP.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=SKELTAL MUSCLE, AND HEART;
 RA Faulkner G., Pallavicini A., Portentun E., Comelli A., Ievoliella C.,
 Trevisan S., Bortolotto G., Scannapieco P., Salamon M., Mouty G.,
 Valle G., Lanfranchi G.;
 RT "ZASP: a new 2-band alternatively spliced PDZ-motif protein.";
 RL J. Cell Biol. 0:0-0(0).
 DR EMBL; AJ133766; CAB46727.1; -
 DR InterPro; IPR001478; PDZ.
 DR Pfam; PF00595; PDZ; 1.
 DR PROSITE; PS50106; PDZ; 1.
 DR SMART; SM00228; PDZ; 1.
 FT NON_TER 1
 SQ SEQUENCE 283 AA; 30791 MW; F6384C88E71767E9 CRC64;

Query Match 12.7%; Score 131; DB 4; Length 283;
 Best Local Similarity 30.1%; Pred. No. 0.00042;
 Matches 46; Conservative 25; Mismatches 74; Indels 8; Gaps 5;

QY 32 LBSGAWGTLLKGLGEHGEPLISYVEEGKADTLLSKLQAGDEVVHINEV-TLSSSRKE 90
 Db 7 LSGPGWGRLOGGKDFNPLTISRTPSKA--AOSQSGGLVVAIDGVNDTWTLE 64
 QY 91 AVSLVKGSKYKTLRVVSLSPPTVSLFDP--OHPQMPDPTRTSFVSSTADGRHEMSCR 149
 Db 65 AOKKITSASVNLSTLQKSKRPISITAPPVOTPLPVIPIHOKVVVN-SPANNDYGERFN 123
 QY 150 PPVVK--WMSPPRTWAARWPKGCIYPROHNT 179

[illegible]

Fri Feb 1 08:56:27 2002

us-09-641-831-2.rspt

Page 7

Db 66 AONKIKGCTGSLNMTLQASAPRPEPVQ-KGEPRKVVKVPVITSPAVSKVTSTNMA 124
QY 146 WSCR-PWTKWSPRPT 161
Db 125 YNKAPRPPGVSVPKVT 141

Search completed: January 31, 2002, 21:05:34
Job time: 97 sec

XX Novel human polynucleotides isolated from human mammary gland cDNA
 PT library, encodes novel human proteins which are useful in diagnosis,
 PT drug screening, clinical trial monitoring or treating behavioural
 PT disorders -
 XX
 PS Claim 3; Page 25; 29pp; English.
 XX
 CC The sequence encodes a novel human protein, NHP#2, which shares
 CC structural motifs with human APXL protein. Nucleotide constructs
 CC encoding functional NHPs are used in gene therapy approaches for the
 CC modulation of NHP expression. NHP oligonucleotides can be used as
 CC hybridisation probes for screening libraries and assessing NHP gene
 CC expression patterns. Also, labeled NHP nucleotide probes can be used to
 CC screen a human genomic library. The NHP nucleotide sequences are also
 CC useful in drug screening techniques for treating symptomatic or
 CC phenotypic manifestations of perturbing the normal function of NHP in the
 CC body. Examples of such manifestations may include obesity and high blood
 CC pressure.
 XX
 SQ Sequence 327 BP; 91 A; 76 C; 96 G; 64 T; 0 other;

Query Match 100.0%; Score 327; DB 22; Length 327;
 Best Local Similarity 100.0%; Pred. No. 1.3e-95;
 Matches 327; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 atgatgagacacactgaagactccacaagccttagtgccacattaaacttaaacacgccc 60
 DB 1 atgatgagacacactgaagactccacaagccttagtgccacattaaacttaaacacgccc 60
 QY 61 accaaggaaggtacattattctgagcattctcctgagggaggagctccctggggtttt 120
 DB 61 accaaggaaggtacattattctgagcattctcctgagggaggagctccctggggtttt 120
 QY 121 actctaaagggtgctgagcagcagagacatttaattctctcaagtcgaagaagg 180
 DB 121 actctaaagggtgctgagcagcagagacatttaattctctcaagtcgaagaagg 180
 QY 181 ggcaagcagacacccctgagctccaaactgcaggctggggtgaggtgtgacatcaat 240
 DB 181 ggcaagcagacacccctgagctccaaactgcaggctggggtgaggtgtgacatcaat 240
 QY 241 gaggtgactctgagcagctccagaagaggcagcttccctgggtgaaggatcctacaag 300
 DB 241 gaggtgactctgagcagctccagaagaggcagcttccctgggtgaaggatcctacaag 300
 QY 301 accctcaggctgtagtgcagttga 327
 DB 301 accctcaggctgtagtgcagttga 327

RESULT 2
 AAS00041
 ID AAS00041 standard; cDNA; 402 BP.
 XX
 AC AAS00041;
 XX
 DT 11-MAY-2001 (first entry)
 XX
 DE Human cDNA encoding novel human protein, NHP#3.
 XX
 KW Human: novel human protein; NHP#3; gene therapy; drug screening;
 KW obesity; high blood pressure; ss.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 1..402
 FT /tag= a
 FT /product= "NHP#3"
 XX
 PN WO200114422-A1.

XX 01-MAR-2001.
 XX
 PF 18-AUG-2000; 2000WO-US22815.
 XX
 PR 24-AUG-1999; 99US-0150511.
 XX
 PA (LEXI-) LEXICON GENETICS INC.
 XX
 PI Turner CA, Zambrowicz B, Friedrich G, Nehls M, Sands AT;
 XX
 DR WPI; 2001-218430/22.
 DR P-PSDB; AAU00033.
 XX
 PT Novel human polynucleotides isolated from human mammary gland cDNA
 PT library, encodes novel human proteins which are useful in diagnosis,
 PT drug screening, clinical trial monitoring or treating behavioural
 PT disorders -
 XX
 PS Claim 5; Page 25; 29pp; English.
 XX
 CC The sequence encodes a novel human protein, NHP#3, which shares
 CC structural motifs with human APXL protein. Nucleotide constructs
 CC encoding functional NHPs are used in gene therapy approaches for the
 CC modulation of NHP expression. NHP oligonucleotides can be used as
 CC hybridisation probes for screening libraries and assessing NHP gene
 CC expression patterns. Also, labeled NHP nucleotide probes can be used to
 CC screen a human genomic library. The NHP nucleotide sequences are also
 CC useful in drug screening techniques for treating symptomatic or
 CC phenotypic manifestations of perturbing the normal function of NHP in the
 CC body. Examples of such manifestations may include obesity and high blood
 CC pressure.
 XX
 SQ Sequence 402 BP; 110 A; 91 C; 117 G; 84 T; 0 other;
 Query Match 98.8%; Score 323; DB 22; Length 402;
 Best Local Similarity 100.0%; Pred. No. 2.7e-94;
 Matches 323; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 atgatgagacacactgaagactccacaagccttagtgccacattaaacttaaacacgccc 60
 DB 1 atgatgagacacactgaagactccacaagccttagtgccacattaaacttaaacacgccc 60
 QY 61 accaaggaaggtacattattctgagcattctcctgagggaggagctccctggggtttt 120
 DB 61 accaaggaaggtacattattctgagcattctcctgagggaggagctccctggggtttt 120
 QY 121 actctaaagggtgctgagcagcagagacatttaattctctcaagtcgaagaagg 180
 DB 121 actctaaagggtgctgagcagcagagacatttaattctctcaagtcgaagaagg 180
 QY 181 ggcaagcagacacccctgagctccaaactgcaggctggggtgaggtgtgacatcaat 240
 DB 181 ggcaagcagacacccctgagctccaaactgcaggctggggtgaggtgtgacatcaat 240
 QY 241 gaggtgactctgagcagctccagaagaggcagcttccctgggtgaaggatcctacaag 300
 DB 241 gaggtgactctgagcagctccagaagaggcagcttccctgggtgaaggatcctacaag 300
 QY 301 accctcaggctgtagtgcgag 323
 DB 301 accctcaggctgtagtgcgag 323
 RESULT 3
 AAS00039
 ID AAS00039 standard; cDNA; 573 BP.
 XX
 AC AAS00039;
 XX
 DT 11-MAY-2001 (first entry)
 XX

```

DE Human cDNA encoding novel human protein, NHP#1.
XX
XX Human; novel human protein; NHP#1, gene therapy; drug screening;
KW obesity; high blood pressure; ss.
XX
XX Homo sapiens.
OS
XX Key Location/Qualifiers
FH 1..573
FT CDS /tag= a
FT /product= "NHP#1"
XX
XX WO200114422-A1.
XX
XX 01-MAR-2001.
XX
XX 18-AUG-2000; 2000WO-US22815.
XX
XX 24-AUG-1999; 99US-0150511.
XX
XX (LEXI-) LEXICON GENETICS INC.
XX
XX Turner CA, Zambowicz B, Friedrich G, Nehls M, Sands AT;
XX
XX WPI; 2001-218430/22.
XX
XX P-PSDB; AAU000031.
XX
XX Novel human polynucleotides isolated from human mammary gland cDNA
PT library, encodes novel human proteins which are useful in diagnosis,
PT drug screening, clinical trial monitoring or treating behavioural
PT disorders -
XX
XX Claim 1; Page 24; 29pp; English.
XX
XX The sequence encodes a novel human protein, NHP#1, which shares
XX structural motifs with human APLX protein. Nucleotide constructs
CC encoding functional NHPs are used in gene therapy approaches for the
CC modulation of NHP expression. NHP oligonucleotides can be used as
CC hybridisation probes for screening libraries and assessing NHP gene
CC expression patterns. Also, labeled NHP nucleotide probes can be used to
CC screen a human genomic library. The NHP nucleotide sequences are also
CC useful in drug screening techniques for treating symptomatic or
CC phenotypic manifestations of perturbing the normal function of NHP in the
CC body. Examples of such manifestations may include obesity and high blood
CC pressure.
CC
XX Sequence 573 BP; 151 A; 152 C; 152 G; 118 T; 0 other:
SQ

Query Match          98.8%; Score 323; DB 22; Length 573;
Best Local Similarity 100.0%; Pred. No. 3,1e-94;
Matches 323; Conservative 0; Mismatches 0; Indels 0; Gaps 0

QY      1 atgataaggacccactgaagacttcacaaagcctatgtgccaatcaaaccttaacagggc 60
       |||||||
Db      1 atgatgagaccactgaagacttcacaaagcctatgtgccaatcaaaccttaacagggc 60

QY      61 accaagaggaaagtaaccttttctcggaagcattcctcgagaagagaagctcccttggttt 120
       |||||||
Db      61 accaagaggaaagtacaatttatctcgagagcatccctgagaagagaagctcccttggttt 120

QY      121 accttaagggtgcctcgagacacgagaaacccaattatcatctctaaggttcgaagaagg 180
       |||||||
Db      121 accttaagggtgcctcgagacacgagaaacccaattatcatctctaaggttcgaagaagg 180

QY      181 ggcacaagcacaccctcagctccaacactcgaggctgggagtgaggtgtgtcacatcaat 240
       |||||||
Db      181 ggcacaagcacaccctcagctccaacactcgaggctgggagtgaggtgtgtcacatcaat 240

QY      241 gaagtgactcttgacactccagaagaagaggaattcccctcgfgaagaagatcctacaag 300
       |||||||
Db      241 gaagtgactcttgacactccagaagaagaggaattcccctcgfgaagaagatcctacaag 300

```

Oy	301	accctcagcctgtagtgcgcac	323
Db	301	accctcagcctgtagtgcgcac	323

RESULT 4			
ID	AA137914		
XX	AA137914	standard: DNA; 462 BP.	
XX	AA137914;		
DT	17-OCT-2001	(first entry)	
DE	Probe #6600 used to measure gene expression in human placenta sample.		
XX	Probe: microarray; human; placenta; antenatal diagnosis;		
KM	genetic disorder; ss.		
XX	Homo sapiens.		
OS	WO200157272-A2.		
PN	09-AUG-2001.		
XX	09-AUG-2001.		
PD	30-JAN-2001; 2001WO-US000663.		
XX	04-FEB-2000; 2000US-0180312.		
XX	26-MAY-2000; 2000US-0207456.		
PR	30-JUN-2000; 2000US-0608408.		
PR	03-AUG-2000; 2000US-0632366.		
PR	21-SEP-2000; 2000US-0234687.		
PR	27-SEP-2000; 2000US-0236359.		
PR	04-OCT-2000; 2000GB-0024263.		
XX	(MOLE-) MOLECULAR DYNAMICS INC.		
PA	Penn SG, Hanzel DK, Chen W, Rank DR;		
PI	WPI; 2001-488897/53.		
XX			
DR			
XX			
PT	Human genome-derived single exon nucleic acid probes useful for		
PT	analyzing gene expression in human placenta -		
XX			
PS	Claim 25; SEQ ID No 6600; 654bp; English.		
XX			
CC	The present invention relates to single exon nucleic acid probes (SENPs).		
CC	The present sequence is one such probe. The probes are useful for		
CC	producing a microarray for predicting, measuring and displaying gene		
CC	expression in samples derived from human placenta. The probes are useful		
CC	for antenatal diagnosis of human genetic disorders.		
XX			
SQ	Sequence 462 BP; 109 A; 102 C; 135 G; 116 T; 0 other;		

Query Match	19.8%;	Score 64.6;	DB 22;	Length 462;
Best Local Similarity	64.2%;	Pred. No. 4.4e-11;		
Matches 97;	Conservative 0;	Mismatches 54;	Indels 0;	Gaps 0;

Oy	175	gaaggaggcaagacagacacctgagctccaaactcagctcggggataagttgtgcac	234
Db	312	gaagagggcagtaaacgcgcgcgtgacaaagtactgtctgagatgagatcgctgcgc	371
Oy	235	ataatgaggtgctctctgacagcctccagaagaagagcgagttccctcgtgtaagagatcc	294
Db	372	atcaatgaactgtctcctcagaaggtttagacaggaagcgattgtcctgtgtaagggttcc	431
Oy	295	tacaagacctcagctggtgtagtgcacagt	325
Db	432	cataagacctgagctggtgcgtcacaaggt	462

RESULT 5	
AA151074	

AAI51074	standard; DNA; 114 BP.
AAI51074;	
17-OCT-2001	(first entry)
Probe #19760	used to measure gene expression in human placenta sample.
Probe; microarray;	human; placenta; antenatal diagnosis;
genetic disorder;	ss.
Homo sapiens.	
WO200157272-A2.	
09-AUG-2001.	
30-JAN-2001;	2001WO-US00663.
04-FEB-2000;	2000US-0180312.
26-MAY-2000;	2000US-0207456.
30-JUN-2000;	2000US-0608408.
03-AUG-2000;	2000US-0632366.
21-SEP-2000;	2000US-0234687.
27-SEP-2000;	2000US-0236359.
04-OCT-2000;	2000GB-0024263.
(MOLE-)	MOLECULAR DYNAMICS INC.
Penn SG,	Hanzel DK, Chen W, Rank DR;
WPI;	2001-488897/53.
Human genome-derived	single exon nucleic acid probes useful for
analyzing gene expression	in human placenta -
Claim 25;	SEQ ID No 19760; 654pp; English.
The present invention	relates to single exon nucleic acid probes (SENP).
The present sequence	is one such probe. The probes are useful for
producing a microarray	for predicting, measuring and displaying gene
expression in samples	derived from human placenta. The probes are useful
for antenatal diagnosis	of human genetic disorders.
Sequence 114	BP; 29 A; 23 C; 36 G; 26 T; 0 other;
Query Match	17.1%; Score 55.8; DB 22; Length 114;
Best Local Similarity	70.1%; Pred. No. 1.7e-08;
Matches 75;	Conservative 0; Mismatches 32; Indels 0; Gaps 0;
QY 211	caggctgggagatgaggtgtgcacatcaatgaggtgactctgagcagctccagaaggag 270
Db 2	ctgggtggagatgagatcgcggtacatgacattggtctctcagggtttagacaggaa 61
QY 271	gcagtttccctggtggaaggatctacaagaccctcaggctggtagt 317
Db 62	gcgatttgctgtggaagggtcccataaagaccctgaagctggtcgt 108
RESULT 6	
AAC98998	
ID AAC98998	standard; cDNA; 1525 BP.
XX AC AAC98998;	
XX 09-MAR-2001	(first entry)
XX Human	pancreatic cancer antigen nucleotide sequence SEQ ID NO:226.
XX Human;	pancreas; pancreatic cancer; pancreatic cancer antigen;
XX detection;	diagnosis; identification; cytostatic; neuroprotective;
XX nontropic;	immunomodulatory; relaxant; contraceptive; gynaecological;

ID AAC63964 standard; DNA; 2566 BP.
 XX
 AC AAC63964;
 XX
 DT 09-FEB-2001 (first entry)
 XX
 DE Methylotrophus methylotrophus diaminopimelate decarboxylase gene.
 XX
 KM Lysine biosynthesis; diaminopimelate decarboxylase;
 KM methanol carbon source; enhanced enzyme activity;
 KM L-lysine acid production; transgenic bacterium; Methylotrophus; ds.
 XX
 OS Methylotrophus methylotrophus.
 XX
 PN WO200061723-A1.
 XX
 PD 19-OCT-2000.
 XX
 PF 07-APR-2000; 2000MO-JF02295.
 XX
 PR 09-APR-1999; 99JP-0103143.
 PR 16-JUN-1999; 99JP-0169447.
 PR 24-DEC-1999; 99JP-0368097.
 XX
 PA (AJIN) AJINOMOTO CO INC.
 PI Gunji Y, Yasueda H, Sugimoto S, Tsujimoto N, Shimaoka M, Miyata Y;
 PI Oda M;
 PI
 DR WPI: 2000-672679/65.
 DR P-PSDB: AAB29520.
 XX
 PT Efficient production of L-amino-acids e.g. L-lysine by fermenting
 PT transformed Methylotrophus bacterium with enhanced dihydrodipicolinate
 PT synthase and aspartokinase activities or casamino acid requirement
 PT using methanol
 XX
 PS Claim 25; Page 77-80; 92pp; Japanese.
 XX
 CC The invention relates to a transformed bacterium of the genus
 CC Methylotrophus which is capable of producing an L-amino acid efficiently
 CC using methanol as the main carbon source. The biosynthetic pathway for
 CC the desired amino acid has enhanced activity in the transformed
 CC bacterium. In particular, the amino acid that is produced is lysine,
 CC and the L-lysine biosynthetic enzymes are not subject to feedback
 CC inhibition by lysine. The invention provides genes (AAC63958-C63964)
 CC encoding the Methylotrophus methylotrophus lysine biosynthetic enzymes
 CC aspartokinase (AAB29516), aspartic semialdehyde dehydrogenase
 CC (AAB29517), dihydrodipicolinate synthase (AAB29518), dihydrodipicolinate
 CC reductase (AAB29519), and diaminopimelate decarboxylase (AAB29520). The
 CC invention also relates to the production of an L-amino acid via culture
 CC of an organism of the invention. The method is useful for the production
 CC of L-amino acids including L-lysine, L-valine, L-leucine, L-isoleucine
 CC or L-threonine. The present sequence represents the Methylotrophus
 CC methylotrophus diaminopimelate decarboxylase gene.
 XX
 SO Sequence 2566 BP; 650 A; 668 C; 631 G; 616 T; 1 other:

Query Match 11.6%; Score 37.8; DB 21; Length 2566;
 Best Local Similarity 58.4%; Pred. No. 0.038;
 Matches 66; Conservative 0; Mismatches 47; Indels 0; Gaps 0;
 Oy 170 tcgaagaaggggcaaacgacacccctgagctcacaactcagctgggagtgagtg 229
 Db 1082 tcaaaagccgcttgaaacgggacatcttcttcaacgtggaacacgtgaatgagctag 1141
 Oy 230 tgcacatcagagtgactcagcagcagcagcagcagcagcagcagcagcagcagc 282
 Db 1142 accgcacacagcagtggtggcgcagcctgggcaaaaagcgcctatttcctcgt 1194

RESULT 8

AAT58627
 ID AAT58627 standard; cDNA; 3984 BP.
 XX
 AC AAT58627;
 XX
 DT 22-MAY-1997 (first entry)
 XX
 DE Protein tyrosine phosphatase cDNA.
 XX
 KM Protein tyrosine phosphatase; PTPH1; focal adhesion;
 KM protein tyrosine kinase; malignancy; cancer; gene therapy;
 KM retrovirus; vector; ss.
 XX
 OS Homo sapiens.
 XX
 PN Key Location/Qualifiers
 PN Key 24..2765
 PN CDS /*tag= a
 XX
 PF US5595911-A.
 XX
 PD 21-JAN-1997.
 XX
 PF 14-MAR-1990; 90US-0494036.
 XX
 PR 01-MAR-1991; 91US-0663579.
 PR 14-MAR-1990; 90US-0494036.
 PR 16-AUG-1993; 93US-0107420.
 XX
 PA (COLD-) COLD SPRING HARBOR LAB.
 XX
 PI Tonks NK;
 PI
 DR WPI: 1997-107583/10.
 DR P-PSDB: AAM12522.
 XX
 PT DNA encoding protein tyrosine phosphatase - for gene therapy of
 PT cancer
 XX
 PS Claim 1; Fig 1A-B; 12pp; English.
 XX
 CC A cDNA clone (AAT58627) codes for a protein tyrosine phosphatase
 CC (PTPH1) (AAM12522) which catalyses the dephosphorylation of proteins
 CC in which tyrosyl residues have been phosphorylated through the
 CC action of a protein tyrosine kinase (PTK). PTPH1 localises to
 CC focal adhesions, a major site of action of oncogenic PTKs. The
 CC cDNA clone was obtd. by amplifying HeLa cDNA using primers based on
 CC conserved segments within the catalytic domains of PTPHs, and using
 CC a PCR product to screen an HeLa cDNA library. The PTPH1 cDNA can
 CC be incorporated into a vector (pref. retroviral) to allow
 CC expression of PTPH1 in mammalian cells in sufficient quantities to
 CC overcome or counteract PTK activity. Phosphorylation of tyrosine
 CC residues at abnormal levels is prevented or reversed, resulting in
 CC the prevention or reversal of malignancy of cells.
 XX
 SO Sequence 3984 BP; 1085 A; 954 C; 931 G; 1014 T; 0 other:

Query Match 11.6%; Score 37.8; DB 18; Length 3984;
 Best Local Similarity 53.8%; Pred. No. 0.046;
 Matches 78; Conservative 0; Mismatches 67; Indels 0; Gaps 0;
 Oy 97 gagggagagcctccctgggttttactcaaaaggtgcttgagcaggaaccatata 156
 Db 1568 gatgaagatggaaaatttggatttaacttaaggaggagtgatcaaaaagatgctctt 1628
 Oy 157 atcacctcgaaggtcgaagaaggggcaaacgacacacccctgagctcacaactcgaagct 216
 Db 1629 gtggtatcaagataaacccagagtcacactcggacacacttgatcccaagctgaagaa 1688
 Oy 217 ggggagtgaggttgacacatcaatg 241
 Db 1689 ggggagtcacatcgttataatcaatg 1713

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 31, 2002, 19:53:37 ; Search time 163.78 Seconds
(without alignments)
1711.719 Million cell updates/sec

Title: US-09-641-831-3

Perfect score: 327

Sequence: 1 atgatgaggaccactgaaga.....ggctgtagtgcgcagttga 327

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N.Geneseq_1101.*

1: /SID82/gcgdata/geneseq/geneseq/NA1980.DAT.*
2: /SID82/gcgdata/geneseq/geneseq/NA1981.DAT.*
3: /SID82/gcgdata/geneseq/geneseq/NA1982.DAT.*
4: /SID82/gcgdata/geneseq/geneseq/NA1983.DAT.*
5: /SID82/gcgdata/geneseq/geneseq/NA1984.DAT.*
6: /SID82/gcgdata/geneseq/geneseq/NA1985.DAT.*
7: /SID82/gcgdata/geneseq/geneseq/NA1986.DAT.*
8: /SID82/gcgdata/geneseq/geneseq/NA1987.DAT.*
9: /SID82/gcgdata/geneseq/geneseq/NA1988.DAT.*
10: /SID82/gcgdata/geneseq/geneseq/NA1989.DAT.*
11: /SID82/gcgdata/geneseq/geneseq/NA1990.DAT.*
12: /SID82/gcgdata/geneseq/geneseq/NA1991.DAT.*
13: /SID82/gcgdata/geneseq/geneseq/NA1992.DAT.*
14: /SID82/gcgdata/geneseq/geneseq/NA1993.DAT.*
15: /SID82/gcgdata/geneseq/geneseq/NA1994.DAT.*
16: /SID82/gcgdata/geneseq/geneseq/NA1995.DAT.*
17: /SID82/gcgdata/geneseq/geneseq/NA1996.DAT.*
18: /SID82/gcgdata/geneseq/geneseq/NA1997.DAT.*
19: /SID82/gcgdata/geneseq/geneseq/NA1998.DAT.*
20: /SID82/gcgdata/geneseq/geneseq/NA1999.DAT.*
21: /SID82/gcgdata/geneseq/geneseq/NA2000.DAT.*
22: /SID82/gcgdata/geneseq/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	327	100.0	327	22	Human cDNA encodin
2	323	98.8	402	22	Human cDNA encodin
3	323	98.8	573	22	Human cDNA encodin
4	64.6	19.8	462	22	Probe #5600 used t
5	55.8	17.1	114	22	Probe #19760 used
6	44.2	13.5	1525	21	Human pancreatic c
7	37.8	11.6	2566	21	Methylphilus meth
8	37.8	11.6	3984	18	Protein tyrosine p
9	37.8	11.5	3984	20	Human FPHI cDNA.
10	36.6	11.2	457	21	Human ORFX ORF1009
11	36.6	11.2	478	20	Human secreted pro

12	36.2	11.1	1696	20	AAX24307	Rat bone mineralis
13	36.2	11.1	1696	21	AAC87708	Rat LIM mineralisa
14	34.8	10.6	325	20	AAV88174	EST clone DU123..
15	34.6	10.6	717	20	AAX24320	Human bone mineral
16	34.6	10.6	717	21	AAC87713	Human LIM minerali
17	34.6	10.6	1456	21	AAC87741	Human LMP-2 (HLMP-
18	34.6	10.6	1575	21	AAC87742	Human LMP-3 (HLMP-
19	34.6	10.6	1620	21	AAC87737	Human LMP expressi
20	34.6	10.6	1644	20	AAX24322	Human bone mineral
21	34.6	10.6	1644	21	AAC87715	Human LIM minerali
22	34.6	10.6	1665	20	AAX24328	Human truncated bo
23	34.6	10.6	1665	21	AAC87738	Human LMP 5'-UTR s
24	34.6	10.6	1689	20	AAX24327	Human bone mineral
25	34.6	10.6	1689	21	AAC87727	Human LIM minerali
26	34	10.4	1511	19	AAV33198	Secreted protein D
27	33.8	10.3	1593	20	AAX242135	Human normal bladd
28	33.4	10.2	1597	17	AAI212908	Human insulin rece
29	33.2	10.2	1336	19	AAV30952	Homo sapiens HLIM-
30	33.2	10.2	1554	21	AAC75976	Human ORFX ORF1531
31	33.2	10.2	1606	21	AAI16133	Human prostate can
32	33.2	10.2	1844	22	AAH18032	Human cDNA sequenc
33	33.2	10.2	2987	9	AAH80456	Sequence of Rhodos
34	33	10.1	376	21	AAC02040	Human secreted pro
35	33	10.1	765	20	AAX24324	Human bone mineral
36	33	10.1	765	21	AAC87726	Human osteosarcoma
37	33	10.1	1619	17	AAI18796	Human immunophilin
38	33	10.1	3109	21	AAC78150	Human cancer assoc
39	32.6	10.0	1444	22	AAE23903	Human secreted pro
40	32.4	9.9	1960	16	AAQ84588	AWML chromosome in
41	32.4	9.9	2680	16	AAQ84590	AWML chromosome in
42	32.4	9.9	2887	16	AAQ84589	AWML chromosome in
43	32.4	9.9	6607	22	AAH98321	Rabbit EST-derived
44	32.4	9.9	31063	22	AAE28533	Genomic fragment #
45	32	9.8	2162	20	AAX40058	Colon cancer assoc

ALIGNMENTS

RESULT 1
ID AAS00040 standard; cDNA; 327 BP.
XX
AC AAS00040;
XX
DT 11-MAY-2001 (first entry)
XX
DE Human cDNA encoding novel human protein, NHP#2.
XX
KW Human; novel human protein; NHP#2; gene therapy; drug screening;
KW obesity; high blood pressure; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 1..327
FT /*tag= a
FT /product= "NHP#2"
XX
PN WO200114422-A1.
XX
PD 01-MAR-2001.
XX
PF 18-AUG-2000; 2000WO-US22815.
XX
PR 24-AUG-1999; 99US-0150511.
XX
PA (LEXI-) LEXICON GENETICS INC.
XX
PI Turner CA, Zambrowicz B, Friedrich G, Nehls M, Sands AT;
XX
DR WPI; 2001-218430/22.
DR P-PSDB; AAU00032.

XX cDNA clone 10-4 encodes a novel rat bone mineralisation protein,
 CC termed RLM (rat LIM mineralisation protein, see AAW97843), that
 CC enhances bone mineralisation in mammalian cells grown in vitro,
 CC and also induces bone formation in vivo when produced in mammals.
 CC Clone 10-4 was isolated from a rat osteosarcoma (UMR 106) cDNA
 CC library using a 260-bp partial sequence (see AAX24308) as probe.
 CC Unlike other known cytokines, RLM is not a secreted protein, but
 CC is instead an intracellular signalling molecule, so can provide
 CC intracellular signalling amplification and easier assessment of
 CC transfected cells. RLM affects mineralisation of the bone matrix
 CC as well as differentiation of cells into the osteoblast lineage.
 CC Suitable clinical applications include enhancement of bone
 CC repair in fractures, bone defects, bone grafting, and normal
 CC homeostasis in patients presenting with osteoporosis. The
 CC invention provides vectors, host cells and methods of inducing
 CC bone formation by transfecting osteogenic precursor cells with
 CC an isolated nucleic acid encoding a human (see AAW97844-45) or
 CC rat LMP. The transfection may occur ex vivo or in vivo by direct
 CC injection of virus or naked plasmid DNA. Also provided are methods
 CC of fusing a spine, inducing systemic bone formation, stimulating
 CC production of osteogenic soluble factor by an osteogenic cell, and
 CC using an antisense oligonucleotide to inhibit LMP expression.
 XX Sequence 1696 BP; 372 A; 557 C; 456 G; 311 T; 0 other;

Query Match 11.1%; Score 36.2; DB 20; Length 1696;
 Best Local Similarity 59.0%; Pred. No. 0.1;
 Matches 62; Conservative 0; Mismatches 43; Indels 0; Gaps 0;
 QY 86 aggcattctctgaggagagctccctgggttttactctaaaggtgctggagcagc 145
 DB ||| || ||||| ||| ||||| || ||||| || ||||| ||| ||| ||| |||
 QY 86 aggtagctgaggagagctccctgggttttactctgaggggcaaggacttca 145
 DB ||| || ||||| ||| ||||| || ||||| || ||||| ||| ||| ||| |||
 QY 146 gagaaccattatctctaaagtcgaagaaggaggcaaacagc 190
 DB ||| || ||||| ||| ||||| || ||||| || ||||| ||| ||| ||| |||
 DB 146 acgtgcccctctccatctctcgtccactcctcgtggaggcaaggccg 190

RESULT 13
 AAC87708
 ID AAC87708 standard; cDNA; 1696 BP.
 AC AAC87708;
 XX
 DT 01-MAR-2001 (first entry)
 XX
 DE Rat LIM mineralisation protein (RLMP) encoding cDNA SEQ ID NO:2.
 XX
 KW LIM mineralisation protein; LMP; bone formation; osteopathic;
 KW osteogenic precursor cell; gene therapy; metabolic bone disease;
 KW osteoporosis; bone degenerative disease; ss.
 XX
 OS Rattus norvegicus.
 XX
 PN WO200006178-A1.
 XX
 PD 09-NOV-2000.
 XX
 PF 28-APR-2000; 2000WO-US11664.
 XX
 PR 30-APR-1999; 99US-0132021.
 XX
 PA (UYEM-) UNIV EMORY.
 XX
 PI Boden SD, Hair GA;
 XX
 XX WPI; 2000-672828/65.
 DR P-FSDB; AAB36471.
 XX
 PT New nucleic acid encoding a human LIM mineralization protein for
 PT inducing or inhibiting bone formation, fusing a spine, stimulating

PT production of an osteogenic cell, or treating bone conditions, such as,
 PT osteoporosis
 XX
 PS Example 12; Page 60; 84pp; English.
 XX
 CC The present invention specifically describes the human LIM
 CC mineralisation proteins (LMP) HLM-2 and HLM-3. LMPs have osteopathic
 CC activity and can be used in gene therapy. LMP nucleic acids can be used
 CC to induce or inhibit bone formation, fuse a spine, stimulate production
 CC of an osteogenic cell, or inhibit the expression of HLM-2 or HLM-3.
 CC They can be used to treat bone conditions, such as, osteoporosis and
 CC other metabolic bone diseases. Antibodies to the LMP proteins encoded
 CC by the nucleic acids are used in marker assays to identify risk factors
 CC in bone degenerative diseases, such as osteoporosis. The nucleic acids
 CC are used in gene therapy for bone formation which leads to the
 CC advantages of: (1) lower production costs; (2) greater efficacy compared
 CC to extracellular treatment regimens due to the ability to achieve
 CC prolonged expression of the intracellular signal; (3) by-passing the
 CC possibility that treatment with extracellular signals might be hampered
 CC due to the presence of limiting numbers or receptors for those signals;
 CC (4) permitting the delivery of transfect potential osteoprogenitor
 CC cells directly to the site where localised bone formation is required;
 CC and (5) permitting systemic bone formation, which provides a treatment
 CC regimen for osteoporosis and other metabolic bone diseases. The present
 CC sequence encodes rat LMP (RLMP), which is used in the exemplification
 CC of the present invention.
 XX
 SQ Sequence 1696 BP; 372 A; 557 C; 456 G; 311 T; 0 other;

Query Match 11.1%; Score 36.2; DB 21; Length 1696;
 Best Local Similarity 59.0%; Pred. No. 0.1;
 Matches 62; Conservative 0; Mismatches 43; Indels 0; Gaps 0;
 QY 86 aggcattctgaggagagctccctgggttttactctaaaggtgctggagcagc 145
 DB ||| || ||||| ||| ||||| || ||||| || ||||| ||| ||| ||| |||
 QY 86 aggtagctgaggagagctccctgggttttactctgcaaggggcaaggacttca 145
 DB ||| || ||||| ||| ||||| || ||||| || ||||| ||| ||| ||| |||
 QY 146 gagaaccattatctctaaagtcgaagaaggaggcaaacagc 190
 DB ||| || ||||| ||| ||||| || ||||| || ||||| ||| ||| ||| |||
 DB 146 acgtgcccctctccatctctcgtccactcctcgtggaggcaaggccg 190

RESULT 14
 AAV88174
 ID AAV88174 standard; cDNA; 325 BP.
 XX
 AC AAV88174;
 XX
 DT 12-FEB-1999 (first entry)
 XX
 DE EST clone DU123.
 XX
 KW Expressed sequence tag; secreted protein; haematopoiesis regulator;
 KW tissue growth; activin; inhibin; tumour invasion suppressor; EST; human;
 KW chemotaxis; chemokinesis; haemostasis; gene therapy; thrombolysis;
 KW receptor; ligand; anti-inflammatory; tumour inhibitor; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO9845437-A2.
 XX
 PD 15-OCT-1998.
 XX
 PF 10-APR-1998; 98WO-US06956.
 XX
 PR 10-APR-1997; 97US-0837312.
 XX
 PA (GEM) GENETICS INST INC.
 XX
 PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D;
 PI Racie LA, Spaulding V, Treacy M;
 XX

DR WPI; 1999-070078/06.

XX New polynucleotides encoding human secreted proteins - derived from
PT e.g. human blood, kidney, foetal lung, placenta, testes, brain,
PT ovary, pituitary, retina and colon cDNA libraries
XX

PS Claim 1; Page 308; 641pp; English.

XX The present sequence represents an expressed sequence tag (EST), and is
CC a polynucleotide of the invention. The polynucleotides of the invention
CC are all secreted EST sequences isolated from a variety of human tissue
CC sources. The EST sequences and proteins encoded by them are predicted to
CC have useful biological activities which would make them suitable for
CC treating, preventing or ameliorating medical conditions in humans and
CC animals, although no supporting data is given. Suggested activities
CC include nutritional activity, immune stimulating or suppressing activity,
CC haemostasis regulating activity, tissue growth activity,
CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, receptor/ligand activity, anti-inflammatory
CC activity, cadherin/tumour invasion suppressor activity, tumour inhibition
CC activity. The EST sequences are also stated to be useful for gene
CC therapy.
XX

SQ Sequence 325 BP; 98 A; 71 C; 85 G; 71 T; 0 other;

Query Match 10.6%; Score 34.8; DB 20; Length 325;
Best Local Similarity 53.7%; Pred. No. 0.15; Mismatches 0; Gaps 0;

Matches 72; Conservative 0; Indels 62; Indels 0; Gaps 0;

QY 55 agggccaccaggaaggatattatctggagcattcctggagggagagcctccg 114

DB 2 aaggaataacagggcagcagggatttctcgcattccatgacgagggcgcccg 61

QY 115 gtttactctaagggtgctgagcagcagacattatcatcctaagggtc 174

DB 62 gggttcagattgcaagggtgcaaggagcagcagccttacaagttgcaagttcga 121

QY 175 gaagggggcaagc 188

DB 122 aatcagagcaagc 135

RESULT 15

AAK24320

AC AAX24320;

DT 07-JUN-1999 (first entry)

DE Human bone mineralisation protein LMP partial cDNA clone.

KW RAMP; LMP; LIM mineralisation protein; bone mineralisation; human;

KW cytokine; spine fusion; fracture repair; bone grafting;

KW osteoporosis; gene therapy; ss.

OS Homo sapiens.

PN WO9906563-A1.

PD 11-FEB-1999.

PF 29-JUL-1998; 98WO-US15814.

PR 02-APR-1998; 98US-0080407.

PR 30-JUL-1997; 97US-0054219.

XX (UYEM-) UNIV EMORY.

XX Boden S, Hair G;

XX WPI; 1999-153793/13.

XX LIM mineralisation proteins - used to induce bone formation from
PT osteogenic precursors, e.g. for treating osteoporosis or repairing
PT fractures
XX

PS Example 17; Page 49; 67pp; English.

XX This is the nucleotide sequence of a human bone mineralisation
CC protein (LMP, or LIM mineralisation protein) partial cDNA clone
CC that was obtained from MG63 osteosarcoma cell cDNA using primers
CC (see AAX24317-18) respectively based on rat LMP cDNA (see AAX24307) and
CC on a previously isolated human partial clone (see AAX24319). A
CC full-length human LMP clone (see AAX24322) encoding a 457-amino acid
CC protein (see AAX27844) was subsequently obtained. LMP enhances bone
CC mineralisation in mammalian cells grown in vitro, and also induces
CC bone formation in vivo. Unlike other cytokines, LMP is not a
CC secreted protein, but is instead an intracellular signalling
CC molecule, so can provide intracellular signalling amplification and
CC easier assessment of transfected cells. Clinical applications
CC include enhancement of bone repair in fractures, bone defects, bone
CC grafting, and normal homeostasis in patients presenting with
CC osteoporosis. A method of inducing bone formation using
CC transfected osteogenic precursor cells is claimed.
XX

SQ Sequence 717 BP; 138 A; 262 C; 219 G; 98 T; 0 other;

Query Match 10.6%; Score 34.6; DB 20; Length 717;
Best Local Similarity 58.1%; Pred. No. 0.24;

Matches 61; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

QY 86 aggcattctggagggagagctcctgggttactctaagggtgcccggagcag 145

DB 14 aggtatgctggaaggccagcacttggtgctccggtcgaaggggcagacttca 73

QY 146 gagaacattatcatcttaagtcgaagaaggggcagaagcag 190

DB 74 atgtccctctccatttcccgctcactcctgggggcaagcgg 118

Search completed: January 31, 2002, 19:53:38
Job time: 191 sec


```

RESULT      1
US-09-124-238A-2
; Sequence 2, Application US/09124238A
; Patent No. 6300127
; GENERAL INFORMATION:
; APPLICANT: Hair, Gregory A.
; APPLICANT: Boden, Scott D.
; TITLE OF INVENTION: No. 6300127el Bone Mineralization Proteins, DNA, Vectors
; FILE REFERENCE: 06148.0115
; CURRENT APPLICATION NUMBER: US/09/124,238A
; CURRENT FILING DATE: 1998-07-29
; PRIOR APPLICATION NUMBER: 60/054,219
; PRIOR FILING DATE: 1997-07-30
; PRIOR APPLICATION NUMBER: 60/080,407
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: MS Word
; SEQ ID NO 2
; LENGTH: 1696
; TYPE: DNA
; ORGANISM: Rattus norvegicus
US-09-124-238A-2

Query Match          11.1%; Score 36.2; DB 4; Length 1696;
Best Local Similarity 59.0%; Pred. No. 0.023;
Matches 62; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY      86 aggcattcctgagaggagagtcctccctgggtttactctaagggttgctgagacg 145
        ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db       86 aggtagctgctgagagagacctgcccttggcgcttcgctcaaggagggaagtcca 145
        ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY      146 gagaaccatatatactctcttaagtcagaagaaggaggaaagcg 190
        || | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db       146 acggccccctctcacctctcgcctcactcctgaggaaggaagcg 190
        || | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT      2
US-09-124-238A-7
; Sequence 7, Application US/09124238A
; Patent No. 6300127
; GENERAL INFORMATION:
; APPLICANT: Hair, Gregory A.
; APPLICANT: Boden, Scott D.
; TITLE OF INVENTION: No. 6300127el Bone Mineralization Proteins, DNA, Vectors
; FILE REFERENCE: 06148.0115
; CURRENT APPLICATION NUMBER: US/09/124,238A
; CURRENT FILING DATE: 1998-07-29
```



```

, PRIOR APPLICATION NUMBER: 60/054,219
, PRIOR FILING DATE: 1997-07-30
, PRIOR APPLICATION NUMBER: 60/080,407
, PRIOR FILING DATE: 1998-04-02
, NUMBER OF SEQ ID NOS: 36
, SOFTWARE: MS Word
, SEQ ID NO 7
, LENGTH: 717
, TYPE: DNA
, ORGANISM: Homo sapiens
US-09-124-238A-7

```

[illegible]

```

RESULT      3
US-09-124-238A-32
, Sequence 32, Application US/09124238A
, Patent No. 6300127
, GENERAL INFORMATION:
, APPLICANT: Hair, Gregory A.
, APPLICANT: Boden, Scott D.
, TITLE OF INVENTION: No. 6300137el Bod
, TITLE OF INVENTION: Expression Syst
, FILE REFERENCE: 06148.0115
, CURRENT APPLICATION NUMBER: US/09/12
, CURRENT FILING DATE: 1998-07-29
, PRIOR APPLICATION NUMBER: 60/054,219
, PRIOR FILING DATE: 1997-07-30
, PRIOR APPLICATION NUMBER: 60/080,407
, PRIOR FILING DATE: 1998-04-02
, NUMBER OF SEQ ID NOS: 36
, SOFTWARE: MS Word
, SEQ ID NO 32
, LENGTH: 1620
, TYPE: DNA
, ORGANISM: Homo sapiens
US-09-124-238A-32

```

[illegible]

RESULT 4
US-09-124-238A-9
Sequence 9, Application US/09124238A
Patent No. 6300127
GENERAL INFORMATION:
APPLICANT: Hair, Gregory A.
APPLICANT: Boden, Scott D.
TITLE OF INVENTION: No. 6300127rel Bone
TITLE OF INVENTION: ExpressionSystems

```

? FILE REFERENCE: 06148.0115
? CURRENT APPLICATION NUMBER: US/09/124,238A
? CURRENT FILING DATE: 1998-07-29
? PRIOR APPLICATION NUMBER: 60/054,219
? PRIOR FILING DATE: 1997-07-30
? PRIOR APPLICATION NUMBER: 60/080,407
? PRIOR FILING DATE: 1998-04-02
? NUMBER OF SEQ ID NOS: 35
? SOFTWARE: MS Word
? SEQ ID NO 9
? LENGTH: 1644
? TYPE: DNA
? ORGANISM: Homo sapiens
? US-09-124-238A-9

```

	Query Match	10.6%	Score 34.6;	DB 4;	Length 1644;
	Best Local Similarity	58.1%;	Pred. NO. 0.076;		
	Matches 61;	Conservative	0;	Mismatches 44;	Indels 0; Gaps 0;
Qy	86	aggcattctcgaggaggagctccctcggggtttactcaagggtggcctggagcag	145		
Db	14	aggtagtctggagggccacgtcggtctccgctgcaagggggcaagacttca	73		
Qy	146	gagacattaatcatcttcaaggctgaaggggggcaagcgag	190		
Db	74	atgfgccctctctcatctccgcgtcactctgagggcgaagcg	118		

```

RESULT      5
US-09-124-238A-33
; Sequence 33, Application US/09124238A
; Patent No. 6300127
; GENERAL INFORMATION:
; APPLICANT: Hair, Gregory A.
; APPLICANT: Boden, Scott D.
; TITLE OF INVENTION: No. 6300127el Body
; TITLE OF INVENTION: Expression Syst
; FILE REFERENCE: 06148.0115
; CURRENT APPLICATION NUMBER: US/09/124
; CURRENT FILING DATE: 1998-07-29
; PRIOR APPLICATION NUMBER: 60/054,219
; PRIOR FILING DATE: 1997-07-30
; PRIOR APPLICATION NUMBER: 60/080,407
; PRIOR FILING DATE: 1998-04-02
; NUMBER OF SEQ ID NOS: 36
; SEQUENCE: MS word
; SEQ ID NO 33
; LENGTH: 1665
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-124-238A-33

```

[illegible]

RESULT 6
US-09-124-238A-22
; Sequence 22, Application US/09124238A
; Patent No. 6300127
; GENERAL INFORMATION:
; APPLICANT: Hair, Gregory A.

```

1  APPLICANT: Boden, Scott D.
2  TITLE OF INVENTION: No. 6300127el Bone Mineralization Proteins, DNA, Vectors
3  TITLE OF INVENTION: Expression Systems
4  TITLE OF INVENTION: Expression Systems
5  FILE REFERENCE: 06148.0115
6  CURRENT APPLICATION NUMBER: US/09/124,238A
7  CURRENT FILING DATE: 1998-07-29
8  PRIOR APPLICATION NUMBER: 60/054,219
9  PRIOR FILING DATE: 1997-07-30
10 PRIOR APPLICATION NUMBER: 60/080,407
11 PRIOR FILING DATE: 1998-04-02
12 NUMBER OF SEQ. ID NOS: 36
13 SOFTWARE: MS Word
14 SEQ ID NO 22
15 LENGTH: 1689
16 TYPE: DNA
17 ORGANISM: Homo sapiens
18 US-09-124-238A-22

```

[illegible]

```

US-08-166-316-1
; Sequence 1, Application US/08166316
; Patent No. 5504192
; GENERAL INFORMATION:
; APPLICANT: Gill, Gordon N.
; APPLICANT: Wu, Rui-Yun
; TITLE OF INVENTION: ENDOCYTIC CODE BINDING PROTEIN
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Spensley Horn Juba & Lubitz
; STREET: 1880 Century Park East, Suite 500
; CITY: Los Angeles
; STATE: California
; COUNTRY: USA
; ZIP: 90067
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/166,316
; FILING DATE: 13-DEC-1993
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Wetherell, Jr., Ph.D., John R.
; REGISTRATION NUMBER: 31,678
; REFERENCE/DOCKET NUMBER: PD-3204
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 455-5110
; TELEFAX: (619) 455-5110
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1597 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
;

```

LOCATION: 1..1362
US-08-166-316-1

	Query Match	10.2%	Score	33.4:	DB	1.	Length	1597:
	Best Local Similarity	55.7%:	Pred No.	0.18:				
Matches	64:	<div style="width:100%; height:1em;"></div>	Conservative	0;	<div style="width:100%; height:1em;"></div>	Mismatches	51;	<div style="width:100%; height:1em;"></div>
						Indels	0;	<div style="width:100%; height:1em;"></div>
						Gaps	0:	<div style="width:100%; height:1em;"></div>
OY	86	aagcaatcctgaaggagaagtccctcggtgatttactctaagtgtygcctggacagc	145					
Dd	14	AAGTATGTCTGGAGGGGCCACCACTTGSGGCTTCGGCTGCAGAAGGGCACAAGATTGA	73					
OY	146	gaaacaataaacatctcttaaggtcgaaaagaygyygcagaacaccctgaq	200					
Dd	74	ATGTGCCCCCTCCCATTTCCCGGCTACTCC TG GGGGCAAAGGGCGCAGCGAG	128					

```

1  RESULT: 8
2  US-08-739-485-6
3  Sequence 6, Application US/08739485
4  Patent No. 5863896
5  GENERAL INFORMATION:
6  APPLICANT: Goli, Surya K.
7  APPLICANT: Hillman, Jennifer L.
8  APPLICANT: Bardman, Olga
9  TITLE OF INVENTION: NOVEL HUMAN LIM P
10 NUMBER OF SEQUENCES: 11
11 CORRESPONDENCE ADDRESS:
12 ADDRESSEE: Incyte Pharmaceuticals,
13 STREET: 3174 Porter Drive
14 City: Palo Alto
15 STATE: CA
16 COUNTRY: US
17 ZIP: 94304
18 COMPUTER READABLE FORM:
19 MEDIUM TYPE: Diskette
20 COMPUTER: IBM Compatible
21 OPERATING SYSTEM: DOS
22 SOFTWARE: FastSEO Version 2.0
23 CURRENT APPLICATION DATA:
24 APPLICATION NUMBER: US/08/739,485
25 FILING DATE:
26 CLASSIFICATION:
27 PRIOR APPLICATION DATA:
28 APPLICATION NUMBER:
29 FILING DATE:
30 ATTORNEY/AGENT INFORMATION:
31 NAME: Billings, Lucy J
32 REGISTRATION NUMBER: 36,749
33 REFERENCE/DOCKET NUMBER: PP-0142 US
34 TELECOMMUNICATION INFORMATION:
35 TELEPHONE: 415-845-0555
36 TELEFAX: 415-845-4166
37 TEXT:
38 INFORMATION FOR SFO ID NO: 6:
39 SEQUENCE CHARACTERISTICS:
40 LENGTH: 1336 base pairs
41 TYPE: nucleic acid
42 STRANDEDNESS: single
43 TOPOLOGY: linear
44 IMMEDIATE SOURCE:
45 LIBRARY: Consensus
46 CLONE: Consensus
47 US-08-739-485-6

```

Query Match	Score	33.2	DB 2	Length	1336
Best Local Similarity	54.0%	Pred.	No. 0.2	58	Indels 0; Gaps 0.
Matches	66	Conservative	0	Mismatches	0

QY	65	aggaagagtcattatctcgtgagcattcctgtgaggaagagctccctcgtggtttacc	124
Db	69	AGGAAGGCGGCGATCCCAAGCAGATGATCTCCCGGGCCCTGAGCCCTTGAGGCTTCAAGC	128

```
QY 125 taaaggtgctgagcagcagagaccattatcatcttaagtcgaagaaggggca 184
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 129 TCTCAGGGGCGATAGACTTCAACGACGCTTTGGTCATCACCAGGATTACACGAGCA 188
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 185 aagcag 190
      ||| |
Db 189 AGCGG 194
      ||| |

RESULT 9
US-09-124-238A-21
; Sequence 21, Application US/09124238A
; Patent No. 6300127
; GENERAL INFORMATION:
; APPLICANT: Hair, Gregory A.
; APPLICANT: Boden, Scott D.
; TITLE OF INVENTION: No. 6300127el Bone Mineralization Proteins, DNA, Vectors,
; TITLE OF INVENTION: Expression Systems
; FILE REFERENCE: 06148-0115
; CURRENT APPLICATION NUMBER: US/09/124,238A
; CURRENT FILING DATE: 1998-07-29
; PRIOR APPLICATION NUMBER: 60/054,219
; PRIOR FILING DATE: 1997-07-30
; PRIOR APPLICATION NUMBER: 60/080,407
; PRIOR FILING DATE: 1998-04-02
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: MS Word
; SEQ ID NO 21
; LENGTH: 765
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-124-238A-21

Query Match 10.1%; Score 33; DB 4; Length 765;
Best Local Similarity 57.1%; Pred. No. 0.18;
Matches 60; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 86 aggcattctgagggaggagctccctgttttactctaaaggtgctgagcag 145
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 74 aggtagtgctgagggcagcacccttccggtcgcagggggaagacttca 133
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 146 gagaaccattatcatcttaagtcgaagaaggggcaagcag 190
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 134 atgtgcctctccattctccggtccacctctctggtgggcaagggc 178
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 10
US-08-533-306A-1
; Sequence 1, Application US/08533306A
; Patent No. 5837457
; GENERAL INFORMATION:
; APPLICANT: Liu, Pu
; APPLICANT: Collins, Francis S.
; APPLICANT: Siciliano, Michael J.
; APPLICANT: Claxton, David
; TITLE OF INVENTION: Markers for Detection of Chromosome 16
; TITLE OF INVENTION: Rearrangements
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Harness, Dickey & Pierce, P.L.C.
; STREET: P.O. Box 828
; CITY: Bloomfield Hills
; STATE: MI
; COUNTRY: USA
; ZIP: 48303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; .CURRENT APPLICATION DATA:
```

```
APPLICATION NUMBER: US/08/533,306A
FILING DATE: September 25, 1995.
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Smith, Deann F.
REGISTRATION NUMBER: 36683
REFERENCE/DOCKET NUMBER: 2115-00869COB
TELECOMMUNICATION INFORMATION:
TELEPHONE: (810) 641-1600
TELEFAX: (810) 641-0270
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1960 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
INDIVIDUAL ISOLATE: Samples 3-6
TISSUE TYPE: Acute myelomonocytic leukemia, M4EO
Tissue Type: subtype (inv16)
POSITION IN GENOME:
CHROMOSOME/SEGMENT: 16(inv(16)(p13q22))
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1731
PUBLICATION INFORMATION:
AUTHORS: Liu, Pu
AUTHORS: Tarle, Susan A.
AUTHORS: Hajra, Amitav
AUTHORS: Claxton, David F.
AUTHORS: Mariton, Paula
AUTHORS: Freedman, Matthew
AUTHORS: Siciliano, Michael J.
AUTHORS: Collins, Francis S.
TITLE: Fusion between transcription factor
TITLE: CBFB/PEBP2B and a myosin heavy chain in acute
TITLE: myelomonocytic leukemia
JOURNAL: Science
DATE: August 1-1993
RELEVANT RESIDUES IN SEQ ID NO: 1: FROM 451 TO 534
PUBLICATION INFORMATION:
AUTHORS: Liu, Pu
AUTHORS: Claxton, David F.
AUTHORS: Mariton, Paula
AUTHORS: Hajra, Jeanette
AUTHORS: Freedman, Matthew
AUTHORS: Chandrasekharappa, Settara C.
AUTHORS: Yanagisawa, Kohsuke
AUTHORS: Stallings, Raymond L.
AUTHORS: Collins, Francis S.
AUTHORS: Siciliano, Michael J.
TITLE: Identification of yeast artificial
TITLE: chromosomes containing the inversion 16 p-arm
TITLE: breakpoint associated with acute myelomonocytic
JOURNAL: Blood
DATE: 1993
US-08-533-306A-1
```

```
Query Match 9.9%; Score 32.4; DB 2; Length 1960;
Best Local Similarity 48.9%; Pred. No. 0.43;
Matches 87; Conservative 0; Mismatches 91; Indels 0; Gaps 0;

QY 124 ctaaaggtggcctgagcagcagagaccattatcatcttaagtcgaagaaggggca 183
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 778 CTGGAAGGGGACCTGNAAGACCTTGAGGCTTCAGGCGGACTCTGCCATCAAGGGGAG 837
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 184 aaagcagacaccctgagctcccaactgcaggtggggatgaggtgtgtgcacatcaatgag 243
```

DB 838 GAAGCCATGACGACGCTACGCAACTGACGCTCAGATGAGAGACTTCAAGAGAGACTG 897
QY 244 gtgaacttgagcagctccagaagaagcagttccctgtygaagaagatccataaga 301
DB 898 GAAGATGCCGCTGCTCCAGATGAGATCTTTGTCACAGCCAAAGAGATGAGAGAGA 955

RESULT 11
US-08-742-923A-1
Sequence 1, Application US/08742923A
Patent No. 5869611
GENERAL INFORMATION:
APPLICANT: Liu, Pu
APPLICANT: Collins, Francis S.
APPLICANT: Siciliano, Michael J.
APPLICANT: Claxton, David
TITLE OF INVENTION: Markers for Detection of Chromosome 16
TITLE OF INVENTION: Rearrangements
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Harness, Dickey & Pierce, P.L.C.
STREET: P.O. Box 828
CITY: Bloomfield Hills
STATE: MI
COUNTRY: USA
ZIP: 48303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/742,923A
FILING DATE: No. 5869611 member 1, 1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Smith, Deann F.
REGISTRATION NUMBER: 36683
REFERENCE/DOCKET NUMBER: 2115-00869DVC
TELECOMMUNICATION INFORMATION:
TELEPHONE: (810) 641-1600
TELEFAX: (810) 641-0270
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1960 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
INDIVIDUAL ISOLATE: Samples 3-6
TISSUE TYPE: Acute myelomonocytic leukemia, M4EO
TISSUE TYPE: subtype (inv16)
POSITION IN GENOME:
CHROMOSOME/SEGMENT: 16[inv(16)(p13q22)]
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1731
PUBLICATION INFORMATION:
AUTHORS: Liu, Pu
AUTHORS: Tarle, Susan A.
AUTHORS: Hajira, Amltay
AUTHORS: Claxton, David F.
AUTHORS: Marlton, Paula
AUTHORS: Freedman, Matthew
AUTHORS: Yanagisawa, Kohsuke
AUTHORS: Stallings, Raymond L.
AUTHORS: Collins, Francis S.
AUTHORS: Siciliano, Michael J.
TITLE: Fusion between transcription factor
TITLE: CBFB/PEBP2B and a myosin heavy chain in acute

TITLE: myelomonocytic leukemia
JOURNAL: Science
DATE: August 1-1993
RELEVANT RESIDUES IN SEQ ID NO: 1: FROM 451 TO 534
PUBLICATION INFORMATION:
AUTHORS: Liu, Pu
AUTHORS: Claxton, David F.
AUTHORS: Marlton, Paula
AUTHORS: Hajira, Jeanette
AUTHORS: Freedman, Matthew
AUTHORS: Chandrasekharappa, Setlara C.
AUTHORS: Yanagisawa, Kohsuke
AUTHORS: Stallings, Raymond L.
AUTHORS: Collins, Francis S.
AUTHORS: Siciliano, Michael J.
TITLE: Identification of yeast artificial
TITLE: chromosomes containing the inversion 16 p-arm
TITLE: breakpoint associated with acute myelomonocytic
TITLE: leukemia
JOURNAL: Blood
DATE: 1993
US-08-742-923A-1

Query Match 9.9%; Score 32.4; DB 2; Length 1960;
Best Local Similarity 48.9%; Pred. No. 0.43;
Matches 87; Conservative 0; Mismatches 91; Indels 0; Gaps 0;

QY 124 ctaaaggtgagcctgagcagcaggaaccattatcatctctaagtgatgaagaaggagc 183
DB 778 CTGAGAGCGGACCTGAAGAGCTGAGGCTTGACGCGGCTGCCATGAAGGAGGAG 837
QY 164 aaagcagcaccctgagctccaaactgagcctgagcctgagcctgagcctgagcctgag 243
DB 838 GAAGCCATGACGACGCTACGCAACTGACGCTCAGATGAGAGACTTCAAGAGAGACTG 897
QY 244 gtgaacttgagcagctccagaagaagcagttccctgtygaagaagatccataaga 301
DB 898 GAAGATGCCGCTGCTCCAGATGAGATCTTTGTCACAGCCAAAGAGATGAGAGAGA 955

RESULT 12
US-08-533-306A-5
Sequence 5, Application US/08533306A
Patent No. 5837457
GENERAL INFORMATION:
APPLICANT: Liu, Pu
APPLICANT: Collins, Francis S.
APPLICANT: Siciliano, Michael J.
APPLICANT: Claxton, David
TITLE OF INVENTION: Markers for Detection of Chromosome 16
TITLE OF INVENTION: Rearrangements
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Harness, Dickey & Pierce, P.L.C.
STREET: P.O. Box 828
CITY: Bloomfield Hills
STATE: MI
COUNTRY: USA
ZIP: 48303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/533,306A
FILING DATE: September 25, 1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Smith, Deann F.
REGISTRATION NUMBER: 36683
REFERENCE/DOCKET NUMBER: 2115-00869COB

```

: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (810) 641-1600
: TELEFAX: (810) 641-0270
: INFORMATION FOR SEQ ID NO: 5:
: SEQUENCE CHARACTERISTICS:
:   LENGTH: 2680 base pairs
:   TYPE: nucleic acid
:   STRANDEDNESS: double
:   TOPOLOGY: linear
: MOLECULE TYPE: cDNA to mRNA
:   HYPOTHETICAL: NO
:   ANTI-SENSE: NO
: ORIGINAL SOURCE:
:   ORGANISM: Homo sapiens
:   INDIVIDUAL ISOLATE: Sample
:   TISSUE TYPE: Acute myelomon
:   TISSUE TYPE: subtype (inv16
:   POSITION IN GENOME:
:   CHROMOSOME/SEGMENT: 16(inv(
:   FEATURE:
:   NAME/KEY: CDS
:   LOCATION: 1..2451
: US-08-533-306A-5

```

	Query Match	9.9%;	Score 32.4;	DB 2:	Length 2680;
	Best Local Similarity	48.9%;	Pred. No. 0.49;		
	Matches 87;	Conservative 0;	Mismatches 91;	Indels 0;	Gaps
QY 124	ctaaagggtggcctgagcagcaggaacatttaattcatctctaaaggtcgaagaagggggc	183			
Db 1498	CTGAGGGGACCTGAAGACCTGTGAGCTTCAGGCCGACTCTGCATCAAGGGAGGGAG	157			
QY 184	aaagcagacacccctgagctccaaactgcaggctggggatgaggtgtgcacatcaatgag	243			
Db 1558	GAAGCCATCAAGCAGCTAGCGAAACTGCAGGCTCAGATGAGGAGCTTTC AAGAGAGCTG	1617			
QY 244	gtgactctgagcagctccgaagaaggagcgagtttccctggtgaaaggatcctacaaga	301			
Db 1618	GAAAGATGCCCGTGCTCCAGAGATCAGATCTTTGCCACAGCCAAAGACAATGAGA	1675			

```

RESULT 13
US-08-742-923A-5
; Sequence 5, Application US/08742923A
; Patent No. 5869611
; GENERAL INFORMATION:
; APPLICANT: Liu, Pu
; APPLICANT: Collins, Francis S.
; APPLICANT: Siciliano, Michael J.
; APPLICANT: Claxton, David
; TITLE OF INVENTION: Markers for Detection of Chromosome 16
; TITLE OF INVENTION: Rearrangements
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Harness, Dickey & Pierce, P.L.C.
; STREET: P.O. Box 828
; CITY: Bloomfield Hills
; STATE: MI
; COUNTRY: USA
; ZIP: 48303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/742,923A
; FILING DATE: NO. 5869611ember 1, 1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Smith, Deann F.
; REGISTRATION NUMBER: 36683

```

```

: REFERENCE/DOCKET NUMBER: 2115-00869DVC
:
: TELECOMMUNICATION INFORMATION:
:
: TELEPHONE: (810) 641-1600
:
: TELEFAX: (810) 641-0270
:
: INFORMATION FOR SEQ ID NO: 5:
:
: SEQUENCE CHARACTERISTICS:
:
: LENGTH: 2680 base pairs
:
: TYPE: nucleic acid
:
: STRANDEDNESS: double
:
: TOPOLOGY: linear
:
: MOLECULE TYPE: cDNA to mRNA
:
: HYPOTHETICAL: NO
:
: ANTI-SENSE: NO
:
: ORIGINAL SOURCE:
:
: ORGANISM: Homo sapiens
:
: INDIVIDUAL ISOLATE: Sample 2
:
: TISSUE TYPE: Acute myelomonocytic leukemia
:
: TISSUE TYPE: subtype (inv16)
:
: POSITION IN GENOME:
:
: CHROMOSOME/SEGMENT: 16[inv(16)(p13q22)]
:
: FEATURE:
:
: NAME/KEY: CDS
:
: LOCATION: 1..2451
:
: US-08-742-923A-5

```

	Query Match	9.9%	Score 32.4;	DB 2;	Length 2680;
	Best Local Similarity	48.9%;	Pred. No. 0.49;		
Matches	87; Conservative	0;	Mismatches 91;	Indels 0;	Gaps 0;
Qy	124	ctaaagggtggcctggagcacggaaccattaatcatcttaaggtcgagaagggggc	183		
Db	1498	CTGAAGGGGACCTTGAAGACTCGGAGCTTCAGGCCACTCTGCATCAAGGGGAGGGAG	1557		
Qy	184	aagcagacaccccttgactccaactcaggctgggggatgaggttgcacatcaatgag	243		
Db	1558	GMAGCCATCAGCAGCTACGCAACTCGAGCTCAGATGAAGGACTTTCAAGAGAGCTG	1617		
Qy	244	gtgactctgagcagctccagaaaggaggcagtttccctggtgaaaggatcctaacaag	301		
Db	1618	GAGATGCCCGTGCTCCAGAGATGTAGATCTTTGCCACAGCCAAGAAGATGAAAGA	1675		

RESULT 14
 US-08-533-306A-3
 ; Sequence 3, Application US/08533306A
 ; Patent No. 5837457
 ; GENERAL INFORMATION:
 ; APPLICANT: Liu, Pu
 ; APPLICANT: Collins, Francis S.
 ; APPLICANT: Siciliano, Michael J.
 ; APPLICANT: Claxton, David
 ; TITLE OF INVENTION: Markers for Detection of Chromosome 16
 ; TITLE OF INVENTION: Rearrangements
 ; NUMBER OF SEQUENCES: 14
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Harness, Dickey & Pierce, P.L.C.
 ; STREET: P. O. Box 828
 ; CITY: Bloomfield Hills
 ; STATE: MI
 ; COUNTRY: USA
 ; ZIP: 48303
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/533.306A
 ; FILING DATE: September 25, 1995
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Smith, Deann F.

GenCore version 4.5
Copyright (c) 1993 - 2000 Comugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 31, 2002, 19:53:38 ; Search time 163.78 Seconds

(Without alignments)
2104.315 Million cell updates/sec

Title: US-09-641-831-5
Perfect score: 402
Sequence: 1 atgatgagaccactgaaga.....atcatgcactgcagccttga 402

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: /SID52/gcgdata/geneseq/geneseq/NA1980.DAT:*
2: /SID52/gcgdata/geneseq/geneseq/NA1981.DAT:*
3: /SID52/gcgdata/geneseq/geneseq/NA1982.DAT:*
4: /SID52/gcgdata/geneseq/geneseq/NA1983.DAT:*
5: /SID52/gcgdata/geneseq/geneseq/NA1984.DAT:*
6: /SID52/gcgdata/geneseq/geneseq/NA1985.DAT:*
7: /SID52/gcgdata/geneseq/geneseq/NA1986.DAT:*
8: /SID52/gcgdata/geneseq/geneseq/NA1987.DAT:*
9: /SID52/gcgdata/geneseq/geneseq/NA1988.DAT:*
10: /SID52/gcgdata/geneseq/geneseq/NA1989.DAT:*
11: /SID52/gcgdata/geneseq/geneseq/NA1990.DAT:*
12: /SID52/gcgdata/geneseq/geneseq/NA1991.DAT:*
13: /SID52/gcgdata/geneseq/geneseq/NA1992.DAT:*
14: /SID52/gcgdata/geneseq/geneseq/NA1993.DAT:*
15: /SID52/gcgdata/geneseq/geneseq/NA1994.DAT:*
16: /SID52/gcgdata/geneseq/geneseq/NA1995.DAT:*
17: /SID52/gcgdata/geneseq/geneseq/NA1996.DAT:*
18: /SID52/gcgdata/geneseq/geneseq/NA1997.DAT:*
19: /SID52/gcgdata/geneseq/geneseq/NA1998.DAT:*
20: /SID52/gcgdata/geneseq/geneseq/NA1999.DAT:*
21: /SID52/gcgdata/geneseq/geneseq/NA2000.DAT:*
22: /SID52/gcgdata/geneseq/geneseq/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	402	100.0	402	22 AAS00041	Human cDNA encoding
2	323	80.3	327	22 AAS00040	Human cDNA encoding
3	323	80.3	573	22 AAS00039	Human cDNA encoding
4	64.6	16.1	462	22 AAI37914	Probe #6600 used t
5	59.2	14.7	32134	22 AAI63522	Human kidney relat
6	59.2	14.7	32134	22 AAI63523	Human kidney relat
7	56.2	14.0	634	22 AAF93252	cDNA encoding SRT
8	55.8	13.9	114	22 AAI51074	Probe #19760 used
9	53.8	13.4	17131	21 AAZ60888	DNA encoding a hum
10	52.8	13.1	391	21 AAC17932	Human secreted pro
11	52.4	13.0	50000	21 AAA96367	Polymorphic repeat

C 12	52	12.9	608	20 AAX20431	Human secreted pro
C 13	51.6	12.8	1159	22 AAH94464	Human foetal cDNA,
C 14	50.6	12.6	313	16 AAT23935	Human gene signatu
C 15	50.6	12.6	362	22 AAI11740	Probe #1673 for ge
C 16	50.6	12.6	362	22 AAI33047	Probe #1733 used t
C 17	50.6	12.6	362	22 AAI01668	Probe #1659 used t
C 18	49.2	12.2	385	21 AAC16802	Human secreted pro
C 19	49.2	12.2	1314	21 AAC79722	Human secreted pro
C 20	49.2	12.2	72604	20 AAI0752	Genomic sequence o
C 21	49.2	12.2	335	21 AAC30196	Human secreted pro
C 22	48.4	12.0	1669	22 AAH18444	Human cDNA sequenc
C 23	48.4	12.0	87350	18 AAX83003	Human WRN genomic
C 24	48.4	11.9	223	21 AAC24831	Human secreted pro
C 25	47	11.7	406	21 AAC26743	Human secreted pro
C 26	47	11.7	8353	18 AAT89396	Methods for diagno
C 27	47	11.7	44453	20 AAX23519	Human kidney amino
C 28	46.8	11.6	3205	22 AAH73404	Human R33 protein
C 29	46.6	11.6	28994	19 AAV15826	Genomic DNA for in
C 30	46.6	11.6	33780	22 AAH24652	Nucleotide sequenc
C 31	46.2	11.5	1691	22 AAH99051	Human EST-derived
C 32	46.2	11.5	1691	22 AAH99080	Human neuroblastom
C 33	45.8	11.4	3530	22 AAF97847	I-FLICE-2 coding s
C 34	45	11.2	2597	22 AAV44807	Human foetal cDNA,
C 35	44.8	11.1	488	22 AAH94277	Human secreted pro
C 36	44.6	11.1	172	21 AAC16201	Human secreted pro
C 37	44.4	11.0	231	21 AAC19578	3'-Flanking sequen
C 38	44.4	11.0	472	17 AAT12399	Human CTLR-L DNA.
C 39	44.4	11.0	2143	19 AAV61937	Human FLICE-like 1
C 40	44.4	11.0	2143	20 AAZ39040	Human FLN-1 encod1
C 41	44.4	11.0	2188	19 AAV21375	Human GI protein 1
C 42	44.4	11.0	2243	19 AAV52968	Human cDNA sequenc
C 43	44.4	11.0	3530	21 AAH18408	Human purh gene ge
C 44	44.4	11.0	41684	21 AAA28150	Human secreted pro
C 45	44.2	11.0	1089	20 AAX37460	Human secreted pro

ALIGNMENTS

RESULT 1	
AA500041	standard; cDNA: 402 BP.
ID	AA500041
NC	AA500041;
XX	
DT	11-MAY-2001 (first entry)
XX	
DE	Human cDNA encoding novel human protein, NHP#3.
XX	
KW	Human: novel human protein; NHP#3; gene therapy; drug screening;
KW	obesity; high blood pressure; ss.
OS	Homo sapiens.
XX	
FT	Key
FT	CDS
FT	Location/Qualifiers
FT	1..402
FT	/*tag= a
FT	/product= "NHP#3"
PN	WO200114422-A1.
XX	
PD	01-MAR-2001.
XX	
PF	18-AUG-2000; 2000WO-US22815.
XX	
PR	24-AUG-1999; 99US-0150511.
XX	
PA	(LEXI-) LEXICON GENETICS INC.
XX	
PI	Turner CA, Zambrowicz B, Friedrich G, Nehls M, Sands AT;
XX	
DR	WPI; 2001-218430/22.
DR	P-PSDB; AAU00033.

XX Novel human polynucleotides isolated from human mammary gland cDNA
PT library, encodes novel human proteins which are useful in diagnosis,
PT drug screening, clinical trial monitoring or treating behavioural
PT disorders
XX
XX Claim 5; Page 25; 29pp; English.
XX
XX The sequence encodes a novel human protein, NHP#3, which shares
XX structural motifs with human APXL protein. Nucleotide constructs
XX encoding functional NHPs are used in gene therapy approaches for the
XX modulation of NHP expression. NHP oligonucleotides can be used as
XX hybridisation probes for screening libraries and assessing NHP gene
XX expression patterns. Also, labeled NHP nucleotide probes can be used to
XX screen a human genomic library. The NHP nucleotide sequences are also
XX useful in drug screening techniques for treating symptomatic or
XX phenotypic manifestations of perturbing the normal function of NHP in the
XX body. Examples of such manifestations may include obesity and high blood
XX pressure.
XX Sequence 402 BP; 110 A; 91 C; 117 G; 84 T; 0 other;

Query Match 100.0%; Score 402; DB 22; Length 402;
Best Local Similarity 100.0%; Pred. No. 6.3e-120;
Matches 402; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 atgatgagacactgaagactccacagcctagtgccacattaaacttaacacggcc 60
Db 1 atgatgagacactgaagactccacagcctagtgccacattaaacttaacacggcc 60
QY 61 accaagggaaggtacatttctctgaggcattctctgaggagagctccctgggtttt 120
Db 61 accaagggaaggtacatttctctgaggcattctctgaggagagctccctgggtttt 120
QY 121 actctaaagggtgctgagacgagacgacattatctctctaaaggtcgaagaagg 180
Db 121 actctaaagggtgctgagacgagacgacattatctctctaaaggtcgaagaagg 180
QY 181 ggcaagcagacacccctgagctccaaactgcagctgggagtggttgacatcaat 240
Db 181 ggcaagcagacacccctgagctccaaactgcagctgggagtggttgacatcaat 240
QY 241 gaggtagctgagcgtccgagaagaggaggttccctggtgaaaggatcctacaag 300
Db 241 gaggtagctgagcgtccgagaagaggaggttccctggtgaaaggatcctacaag 300
QY 301 accctcaggtgtagtgcgagaaatgggttctgtctgttgcgcagaaatgggaaggta 360
Db 301 accctcaggtgtagtgcgagaaatgggttctgtctgttgcgcagaaatgggaaggta 360
QY 361 gtggtattcatagggatgatcatcatgcactgcagccttga 402
Db 361 gtggtattcatagggatgatcatcatgcactgcagccttga 402

RESULT 2
AAS00040
ID AAS00040 standard; cDNA; 327 BP.
XX AC AAS00040;
XX
XX 11-MAY-2001 (first entry)
XX
XX Human cDNA encoding novel human protein, NHP#2.
XX Human; novel human protein; NHP#2; gene therapy; drug screening;
XX obesity; high blood pressure; ss.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX CDS 1..327
XX

FT /*tag= a
FT /product= "NHP#2"
XX
XX WO200114422-A1.
XX
XX 01-MAR-2001.
XX
XX 18-AUG-2000; 2000WO-US22815.
XX
XX 24-AUG-1999; 99US-0150511.
XX
XX (LEXI-) LEXICON GENETICS INC.
XX
XX Turner CA, Zambrowicz B, Friedrich G, Nehls M, Sands AT;
XX
XX WPI; 2001-218430/22.
XX
XX P-FSDB; AAO00032.
XX
XX Novel human polynucleotides isolated from human mammary gland cDNA
XX library, encodes novel human proteins which are useful in diagnosis,
XX drug screening, clinical trial monitoring or treating behavioural
XX disorders
XX
XX Claim 3; Page 25; 29pp; English.
XX
XX The sequence encodes a novel human protein, NHP#2, which shares
XX structural motifs with human APXL protein. Nucleotide constructs
XX encoding functional NHPs are used in gene therapy approaches for the
XX modulation of NHP expression. NHP oligonucleotides can be used as
XX hybridisation probes for screening libraries and assessing NHP gene
XX expression patterns. Also, labeled NHP nucleotide probes can be used to
XX screen a human genomic library. The NHP nucleotide sequences are also
XX useful in drug screening techniques for treating symptomatic or
XX phenotypic manifestations of perturbing the normal function of NHP in the
XX body. Examples of such manifestations may include obesity and high blood
XX pressure.
XX
XX Sequence 327 BP; 91 A; 76 C; 96 G; 64 T; 0 other;

Query Match 80.3%; Score 323; DB 22; Length 327;
Best Local Similarity 100.0%; Pred. No. 1.8e-94;
Matches 323; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 atgatgagacactgaagactccacagcctagtgccacattaaacttaacacggcc 60
Db 1 atgatgagacactgaagactccacagcctagtgccacattaaacttaacacggcc 60
QY 61 accaagggaaggtacatttctctgaggcattctctgaggagagctccctgggtttt 120
Db 61 accaagggaaggtacatttctctgaggcattctctgaggagagctccctgggtttt 120
QY 121 actctaaagggtgctgagacgagacgacattatctctctaaaggtcgaagaagg 180
Db 121 actctaaagggtgctgagacgagacgacattatctctctaaaggtcgaagaagg 180
QY 181 ggcaagcagacacccctgagctccaaactgcagctgggagtggttgacatcaat 240
Db 181 ggcaagcagacacccctgagctccaaactgcagctgggagtggttgacatcaat 240
QY 241 gaggtagctgagcgtccgagaagaggaggttccctggtgaaaggatcctacaag 300
Db 241 gaggtagctgagcgtccgagaagaggaggttccctggtgaaaggatcctacaag 300
QY 301 accctcaggtgtagtgcgag 323
Db 301 accctcaggtgtagtgcgag 323

RESULT 3
AAS00039
ID AAS00039 standard; cDNA; 573 BP.
XX

RESULT 5
AAI63522/c
ID AAI63522 standard; DNA; 32134 BP.
XX
AC AAI63522;
XX 22-OCT-2001 (first entry)
XX Human kidney related polynucleotide SEQ ID NO 837.
XX
XX Human: kidney antigen; immunosuppressive; antiarthritic; antirheumatic;
KW antiproliferative; cytostatic; cardiant; vasotropic; cerebroprotective;
KW neotropic; neuroprotective; antibacterial; virucide; fungicide;
KW ophthalmological; antiallergic; hepatotropic; antidiabetic;
KW antinflammatory; antitumor; antineoplastic; antiparasitic;
KW gene therapy; cancer; immune disorder; cardiovascular disorder;
KW neurological disease; infection; ds.
XX
OS Homo sapiens.
XX
PN WO200155323-A2.
XX
XX 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01343.
XX
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 05-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234224.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.

FT exon 7292..7384
FT /*tag= e
FT /number= 2
FT intron 7385..13759
FT /*tag= f
FT exon 13760..13830
FT /*tag= g
FT intron 13831..14062
FT /*tag= h
FT exon 14063..15251
FT /*tag= i
FT /number= 4
PN WO200005382-A2.
XX 03-FEB-2000.
XX 23-JUL-1999; 99WO-IB01353.
XX 23-JUL-1998; 98US-0093940.
XX (GEST) GENSET.
XX Bouquellet L;
XX WPI; 2000-182704/16.
XX P-PSDB; AAY68909.
XX New isolated human geranyl-geranyl pyrophosphate synthetase nucleic
XX acids, used to develop agents for the diagnosis of, e.g. pathologies
XX related to a defect in the mevalonic biosynthetic pathway -
XX Claim 1; Page 72-79; 86pp; English.
XX The present sequence represents a genomic sequence of human
XX geranylgeranyl pyrophosphate synthetase (hGGPPS). The sequence
XX comprises the 5' regulatory region, the exons and introns, and
XX 3' regulatory region. Two differently spliced mRNAs exist for this
XX gene. The first spliced mRNA is derived from a CDNA (AAZ60888) which
XX comprises exons 1, 2, 3 and 4. The second mRNA is derived from a
XX CDNA (AAZ60889) which comprises 1b1s, 2, 3, and 4. The hGGPPS gene is
XX located on chromosome 1, at the 1q42-1q43 locus. This chromosome 1
XX locus has been shown to carry a predisposing gene for prostate cancer.
XX The nucleic acids encoding hGGPPS can be used for screening for agents
XX which modulate the expression of the hGGPPS gene. Such agents can be
XX used in therapeutic applications. The biallelic markers associated with
XX the hGGPPS gene can be used for the diagnosis of diseases related to
XX an alteration in the regulatory or coding regions of hGGPPS, such as
XX pathologies related to a defect in the mevalonic biosynthetic pathway.
XX The products can also be used for detection, diagnosis and drug
XX screening.
XX Sequence 17131 BP; 5110 A; 3434 C; 3759 G; 4816 T; 12 other;

Query Match 13.4%; Score 53.8; DB 21; Length 17131;
Best Local Similarity 79.0%; Pred. No. 7.9e-07;
Matches 64; Conservative 0; Mismatches 17; Indels 0; Gaps 0;
QY 322 agaaatgggtctgtatgttcccaagaatggaaggtagtggtatcattcaggcatgat 381
Db 7720 AGAGACGGGCTCACTATGTCGCAAGCTGGAGTGTAGTGTATTCACAGGTGTGAT 7661
QY 382 catcatgcactgcagccttga 402
Db 7660 CATGTGTGCACTATAGCCTTAA 7640
RESULT 10
AAC17932/c

ID AAC17932 standard; cDNA; 391 BP.
XX AAC17932;
XX 06-OCT-2000 (first entry)
XX Human secreted protein 5' EST, SEQ ID NO: 22007.
DE Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping; ss.
KW Homo sapiens.
OS EP1033401-A2.
XX 06-SEP-2000.
XX 21-FEB-2000; 2000EP-0200610.
XX 26-FEB-1999; 99US-0122487.
XX (GEST) GENSET.
XX Dumas Milne Edwards J, Duclert A, Giordano J;
XX WPI; 2000-500381/45.
XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
XX obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
XX diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX Claim 1; SEQ ID 22007; 71pp + CD-ROM; English.
XX The present sequence is one of a large number of 5' ESTs derived from
XX mRNAs encoding secreted proteins. No ORF has yet been conclusively
XX identified within the present sequence. The 5' ESTs were prepared from
XX total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
XX sequences usually correspond mainly to the 3' untranslated region (UTR)
XX of the mRNA because they are often obtained from oligo-dT primed cDNA
XX libraries. Such ESTs are not well suited for isolating cDNA sequences
XX derived from the 5' ends of mRNAs and even in those cases where longer
XX cDNA sequences have been obtained, the full 5' UTR is rarely included.
XX 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
XX used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
XX in diagnostic, forensic, gene therapy and chromosome mapping procedures.
XX They are used to obtain upstream regulatory sequences and to design
XX expression and secretion vectors.
XX Sequence 391 BP; 137 A; 75 C; 94 G; 85 T; 0 other;

Query Match 13.1%; Score 52.8; DB 21; Length 391;
Best Local Similarity 78.8%; Pred. No. 3e-07;
Matches 63; Conservative 0; Mismatches 17; Indels 0; Gaps 0;
QY 323 gaatgggtctgtatgttcccaagaatggaaggtagtggtatcattcaggcatgatc 382
Db 311 GAGATGGAGTTTGTATGTTGCCAGGCTGGAGTGTAGTGTATTCACAGGCGTGATC 252
QY 383 atcatgcactgcagccttga 402
Db 251 ATAGCTCCAGGCAGCCTAGA 232
RESULT 11
AAA96367/c
ID AAA96367 standard; DNA; 50000 BP.
XX AAA96367;
XX 08-FEB-2001 (first entry)
XX Polymorphic repeat microsatellite sequences present in the CTLA4 locus.

XX Autoimmune disease; polymorphic microsateellite repeat; PMR; CD28 gene;
 KW ICOS gene; CTLA4 gene; costimulatory receptor gene locus; CGRL; lupus;
 KW Insulin-dependent diabetes mellitus; IDDM; Addison's disease; leprosy;
 KW Graves disease; autoimmune hypothyroidism; myasthenia gravis; thymoma;
 KW thyroiditis; postpartum thyroiditis; rheumatoid arthritis;
 KW Hashimoto's disease; coeliac disease; ss.
 XX
 OS Homo sapiens.
 XX
 PN W0200056856-A2.
 PD 28-SEP-2000.
 XX
 PF 24-MAR-2000; 2000MO-US07938.
 XX
 PR 25-MAR-1999; 99US-0126215.
 XX
 PA (GEMV) GENETICS INST INC.
 XX
 PI Ling V, Wu P, Gray GS;
 XX
 DR WPI; 2000-628257/60.
 XX
 PT Determining predisposition of humans to develop autoimmune disease
 PT involves detecting polymorphic microsateellite repeat sequence within
 PT human costimulatory receptor gene locus
 XX
 PS Disclosure; Page 128-142; 160pp; English.
 XX
 CC Two human bacterial artificial chromosome (BAC) clones that included
 CC and flanked the human CTLA-4 locus were cloned and sequenced. The
 CC sequence data was assembled into a contiguous sequence that is presented
 CC in AAG96363-68. AAG96363-64 comprise BAC clone 22700, and AAG96365-68
 CC comprise BAC clone 22608. The sequences contain polymorphic
 CC microsatellite repeat (PMR) sequences. The specification describes a
 CC method for determining the predisposition of a human subject to develop
 CC autoimmune disease. The method comprises detecting a PMR sequence in the
 CC CD28, ICOS gene or CTLA4 gene of the human costimulatory receptor gene
 CC locus (hCGR). PMR sequences vary in length among individuals and can be
 CC amplified to generate products that differ in size. These products can
 CC then be detected by rapid and convenient high resolution processes. The
 CC method is useful for determining the predisposition of insulin-dependent
 CC diabetes mellitus (IDDM), Addison's disease, Graves disease, autoimmune
 CC hypothyroidism, myasthenia gravis, thymoma, lupus, thyroiditis,
 CC postpartum thyroiditis, rheumatoid arthritis, Hashimoto's disease,
 CC coeliac disease and leprosy. PMR sequences within hCGR are useful as
 CC markers in a variety of assays and in the field of forensic medicine,
 CC disease diagnosis and human genome mapping.
 CC
 XX
 SQ Sequence 50000 BP; 14612 A; 9948 C; 10072 G; 15368 T; 0 other;

Query Match 13.0%; Score 52.4; DB 21; Length 50000;
 Best Local Similarity 79.5%; Pred. No. 3,6e-06;
 Matches 62; Conservative 0; Mismatches 16; Indels 0; Gaps 0;
 OY 322 agaattgggtctgtctatgtccagaaatgaaagtgtgcatatgagcatgat 381
 DB 18241 AAGATGGGGTCTCACTATGTTGCCCAACTGAGTGCAGTGTCTTTCACAGGTGTGAT 18182
 OY 382 catcatgacatgcagcct 399
 DB 18181 CATAGGCGACTACAGCCT 18164

RESULT 12
 AAX20431/C
 ID AAX20431 standard; DNA; 608 BP.
 AC AAX20431;
 XX
 DT 04-MAY-1999 (first entry)

XX Human secreted protein gene 20.
 XX
 KW Human; secreted protein; fusion protein; gene therapy; protein therapy;
 KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukemia;
 KW developmental abnormality; foetal deficiency; blood; allergy; renal; ds;
 KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
 KW* Inflammation; ischemic shock; Alzheimer's disease; osteoclast; AIDS;
 KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
 KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
 KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
 XX
 OS Homo sapiens.
 XX
 PN W09906423-A1.
 PD 11-FEB-1999.
 XX
 PF 29-JUL-1998; 98MO-US15949.
 XX
 PR 19-AUG-1997; 97US-0056730.
 PR 30-JUL-1997; 97US-0054209.
 PR 30-JUL-1997; 97US-0054211.
 PR 30-JUL-1997; 97US-0054212.
 PR 30-JUL-1997; 97US-0054213.
 PR 30-JUL-1997; 97US-0054214.
 PR 30-JUL-1997; 97US-0054215.
 PR 30-JUL-1997; 97US-0054217.
 PR 30-JUL-1997; 97US-0054218.
 PR 30-JUL-1997; 97US-0054234.
 PR 30-JUL-1997; 97US-0054236.
 PR 18-AUG-1997; 97US-0055968.
 PR 18-AUG-1997; 97US-0055969.
 PR 18-AUG-1997; 97US-0055972.
 PR 19-AUG-1997; 97US-0056534.
 PR 19-AUG-1997; 97US-0056543.
 PR 19-AUG-1997; 97US-0056554.
 PR 19-AUG-1997; 97US-0056561.
 PR 19-AUG-1997; 97US-0056727.
 PR 19-AUG-1997; 97US-0056729.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Carter KC, Endress GA, Fan P, Feng P, Kyaw H, Lafleur DW;
 PI Li Y, Moore PA, Rosen CA, Ruben SM, Shi Y, Wei Y;
 PI Zeng Z;
 XX
 DR WPI; 1999-153691/13.
 XX
 PT P-PSDB; AAY00277.
 PT
 PT New isolated human genes and the secreted polypeptides they encode -
 PT disorders, immune diseases, inflammation or blood disorders
 XX
 PS Claim 1; Page 219; 312pp; English.

This sequence represents a nucleic acid molecule which encodes a
 secreted human protein. The gene number is given in the descriptor line.
 The gene can be used to generate fusion proteins (e.g. AAX20403) for increasing the
 stability of the fused protein as compared to the human protein only.
 The invention relates to 83 novel genes and their fragments (nucleic
 acid sequences: AAX20412-X20499; amino acid sequences: AAY00258-Y00377)
 which are useful for preventing, treating or ameliorating medical
 conditions e.g. by protein or gene therapy. Also, pathological
 conditions can be diagnosed by determining the amount of the new
 polypeptides in a sample or by determining the presence of mutations in
 the new polynucleotides. Specific uses are described for each of the 86
 polynucleotides, based on which tissues they are most highly expressed in
 (see AAX20412 for described uses).

Sequence 608 BP; 162 A; 192 C; 113 G; 138 T; 3 other;


```

RESULT      2
US-09-124-238A-32
/ Sequence 32, Application US/09124238A
/ Patent No. 6300127
/ GENERAL INFORMATION:
/ APPLICANT: Hair, Gregory A.
/ APPLICANT: Boden, Scott D.
/ TITLE OF INVENTION: No. 6300127el Bone Mineralization Proteins, DNA, Vectors,
/ TITLE OF INVENTION: Expression Systems
/ FILE REFERENCE: 06148.0115
/ CURRENT APPLICATION NUMBER: US/09/124,238A
/ CURRENT FILING DATE: 1998-07-29
/ PRIOR APPLICATION NUMBER: 60/054,219
/ PRIOR FILING DATE: 1997-07-30
/ PRIOR APPLICATION NUMBER: 60/080,407
/ PRIOR FILING DATE: 1998-04-02
/ NUMBER OF SEQ ID NOS: 36
/ SOFTWARE: MS word
/ SEQ ID NO 32
/ LENGTH: 1620
/ TYPE: DNA
/ ORGANISM: Homo sapiens
US-09-124-238A-32

```

Query Match	6.5%; Score 37.4; DB 4; Length 1620;
Best Local Similarity	49.8%; Pred. No. 0.016;
Matches 126; Conservative	0; Mismatches 121; Indels 6; Gaps 1;
Qy 86	aggatttcctggaggagagctccctgggttttactctaaagggtgacctgagacagc 145
Db 14	aggtatgctggaggccagacctggggcttccggctcaagggggcaaggactcca 73
Qy 146	gagaaaccattaatctcttaaggctgaagaagggggcaagcagacacctgagctcca 205
Db 74	atgtgccctctccattctccggctcactcctgggggcaagcggc-----cgaggccg 127
Qy 206	aactgcagggtggggatgaggttgtcacatcaatgagtgactctgagcagctccagaa 265
Db 128	gagtgggcgtgggtgactgggtgctgagcatcgatgcggagatgcgggtgagctcacac 187
Qy 266	aggagcaggttccctggtgaaaggatctctacaagacctcaggctggtagtgcgcagcc 325
Db 188	accatcaagctcgaacaagatccggcctcggggagcgcctcagcctgggctcagca 247
Qy 326	tctccccaccggt 338
Db 248	gggcccaagccggt 260

```

RESULT      3
US-09-124-238A-9
; Sequence 9, Application US/09124238A
; Patent No. 6300127
; GENERAL INFORMATION:
; APPLICANT: Hair, Gregory A.
; APPLICANT: Boden, Scott D.
; TITLE OF INVENTION: No. 6300127el Bone Mineralization Proteins, DNA, Vectors,
; TITLE OF INVENTION: Expression Systems
; FILE REFERENCE: 06148.0115
; CURRENT APPLICATION NUMBER: US/09/124,238A
; CURRENT FILING DATE: 1998-07-29
; PRIOR APPLICATION NUMBER: 60/054,219
; PRIOR FILING DATE: 1997-07-30
; PRIOR APPLICATION NUMBER: 60/080,407
; PRIOR FILING DATE: 1998-04-02
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: MS Word
; SEQ ID NO 9
; LENGTH: 1644
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-124-238A-9

```

Query Match	6.5%;	Score 37.4;	DB 4;	Length 1644;
Best Local Similarity	49.8%;	Prod. No. 0.016;		
Matches 126;	Conservative	0;	Mismatches 121;	Indels 6; Gaps 1;
QY	86	aggcattctcgaggaggagctccctgggttttactctaaagggtggcctggagcacg	145	
Db	14	aggtagtctggaggggcagcaccttggggtctcggtcgaaaggggcaaggaattcca	73	
QY	146	gagaaccattaatcatctcttaaggctgaagggggcgaagcagacacccctgagctcca	205	
Db	74	atgtgccctctccattctcccggtcacctctctggggcgaagcggc-----cgaggccg	127	
QY	206	aactcgaggctgggatgaggtgtggcacatcatgagtgactctagcagctccagaa	265	
Db	128	gagtggccgtgggtgactgggtgctgagcatcgtatgcgagaaatgcgggtagcctcacg	187	
QY	266	aggagcgagtttccctgggtgaaggatctctacaagaccctcaggtggtgagtcgcagcc	325	
Db	188	acatcgagctcagaaacaagatccgggctcggggagcgctcagcctgggctcagca	247	
QY	326	tctcccaccagggt	338	
Db	248	gggccagccaggt	260	

RESULT 4
US-09-124-238A-33
; Sequence 33, Application US/09124238A
; Patent No. 6300127
; GENERAL INFORMATION:
; APPLICANT: Hair, Gregory A.
; APPLICANT: Boden, Scott D.
; TITLE OF INVENTION: No. 6300127el Bone Mineralization Proteins, DNA, Vectors,
; TITLE OF INVENTION: Expression Systems
; FILE REFERENCE: 06148.0115
; CURRENT APPLICATION NUMBER: US/09/124,238A
; CURRENT FILING DATE: 1998-07-29
; PRIOR APPLICATION NUMBER: 60/054,219
; PRIOR FILING DATE: 1997-07-30
; PRIOR APPLICATION NUMBER: 60/080,407
; PRIOR FILING DATE: 1998-04-02
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: MS Word
; SEQ ID NO 33
; LENGTH: 1665
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-124-238A-33

Query Match	6.5%	Score 37.4;	DB 4;	Length 1665;
Best Local Similarity	49.8%;	Pred. No. 0.016;		
Matches 126;	Conservative 0;	Mismatches 121;	Indels 6;	Gaps 1;
Qy	86	aggcattctggaggaggagactccctcctgggttttactctaaagggtggcctggagcacg	145	
Db	59	aggtagctggagggccagacactggggtctccgctgcaagggggcaaggacttca	118	
Qy	146	gagaaaccttaatactcttaaggctgaagaagggggcaagcagacacctgagctcca	205	
Db	119	atgtgccctctccattcccggtcactcctggggcgaaggcgc-----gcaggccg	172	
Qy	206	aactgcaggctggggatgaggtgtgtgcacatcaatgactgagctgagcagctcagaa	265	
Db	173	gagtgccgctggggcgactgggtgactgcatgatggcgagaaatcggtgagctcacac	232	
Qy	266	aggaggcagtttccctggtgaaaggatcctacagacctcaggctgggtagtgcgcagcc	325	
Db	233	acatcgaaagtacagacaagatccggcctcggggagcgctcagcctgggcctcagca	292	
Qy	326	tctcccaccgggt	338	

QY 3 gatgaggaccactgaagacttcacaaagcctagtcacattaaacttaacagggccac 62
Db 1380 RRR 1321
QY 63 caagggaaggtacattatctggggcattctggaggagagcctcctgggtttac 122
Db 1320 RRR 1261
QY 123 tctaaagggtgctggagcagcagacacattatctcttaagtcgaaggggg 182
Db 1260 RRR 1201
QY 183 caagcagacacacctgagctccaaactgcagcgtgggatgaggtgtgacatcaatga 242
Db 1200 RRR 1141
QY 243 ggtgactgagcagctccgaaaggagcagtttccctggtgaaagatcctacagac 302
Db 1140 RRR 1081
QY 303 cctcagcgtgtagtcgacgacctccaccgg 337
Db 1080 RRR 1046

RESULT 8
US-09-124-238A-21
; Sequence 21, Application US/09124238A
; Patent No. 6300127
; GENERAL INFORMATION:
; APPLICANT: Hair, Gregory A.
; TITLE OF INVENTION: No. 6300127el Bone Mineralization Proteins, DNA, Vectors,
; FILE OF INVENTION: Expression Systems
; FILE REFERENCE: 06148.0115
; CURRENT APPLICATION NUMBER: US/09/124,238A
; PRIOR FILING DATE: 1998-07-29
; PRIOR APPLICATION NUMBER: 60/054,219
; PRIOR FILING DATE: 1997-07-30
; PRIOR APPLICATION NUMBER: 60/080,407
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: MS Word
; SEQ ID NO 21
; LENGTH: 765
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-124-238A-21

Query Match 6.0%; Score 34.2; DB 4; Length 765;
Best Local Similarity 49.0%; Pred. No. 0.12; Mismatches 123; Indels 6; Gaps 1;
Matches 124; Conservative 0;
QY 86 aggcattctgaggagagcctcctgggttttacttaagggtgctgagcagc 145
Db 74 aggtagtgctgaggagcagcagcctggtctccgctgcaaggggcagacttca 133
QY 146 gagaaccattatctcttaagtcgaagaggagggcgaagcagacacctgagctcca 205
Db 134 atgtgctctctcattctccgctcactctggtgggcaagg-----ccgtgagggccg 187
QY 206 aactgagcgtgggatgaggtgtgacatcaatgaggtgactctgagcagctccagaa 265
Db 188 gagtggccgtaagtgaactggtgctgagcagcagcagcagcagcagcagcagc 247
QY 266 agggagcagtttccctggtgaagagatcctacagaccctcagcgtggtgagcagcc 325
Db 248 acatcgaagcagcaacaagatccggtcgtggggagcgcctcagcgtgggctcaaca 307
QY 326 tctcccccagcgt 338
Db 308 gggccagcagcgt 320

RESULT 9
US-08-967-513-1
; Sequence 1, Application US/08967513
; Patent No. 5783436
; GENERAL INFORMATION:
; APPLICANT: Robert P. Hausinger
; TITLE OF INVENTION: Mutant Urease and Method
; TITLE OF INVENTION: of Use For Determination
; TITLE OF INVENTION: of Urea
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ian C. McLeod
; STREET: 2190 Commons Parkway
; CITY: Okemos
; STATE: Michigan
; COUNTRY: USA
; ZIP: 48864
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 360 Kb
; MEDIUM TYPE: Storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/967,513
; FILING DATE: 11-NOV-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/687,645
; FILING DATE: July 26, 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Ian C. McLeod
; REGISTRATION NUMBER: 20,931
; REFERENCE/DOCKET NUMBER: MSU 4.1-309
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (517) 347-4100
; TELEFAX: (517) 347-4103
; TELEX: No. 5783436e
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2400
; TYPE: nucleotides
; STRANDEDNESS: Single
; TOPOLOGY: Linear
; MOLECULE TYPE:
; DESCRIPTION: cDNA
; HYPOTHEetical: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Klebsiella aerogenes
; STRAIN: CG253
; INDIVIDUAL ISOLATE:
; CELL TYPE: N/A
; FEATURE:
; NAME/KEY: cDNA encoding mutant urease
; NAME/KEY: 'H2190
; LOCATION: Modification at position 1312 to
; LOCATION: glutamine
; IDENTIFICATION METHOD: Sequencing
; OTHER INFORMATION:
US-08-967-513-1

Query Match 5.9%; Score 33.8; DB 1; Length 2400;
Best Local Similarity 46.4%; Pred. No. 0.3;
Matches 110; Conservative 0; Mismatches 127; Indels 0; Gaps 0;
QY 136 ctggagcagcagagacattatctcttaagtcgaagagggggcaagcagacacc 195
Db 127 CCGGAGTCCGTGGCCTGATCAGCGCTTTATTATGAGAGCGCTCGGGACGCAAAAGC 186

IDENTIFICATION METHOD: Sequencing
OTHER INFORMATION:

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 31, 2002, 19:16:22 ; Search time 1599.5 Seconds
(without alignments)
5909.899 Million cell updates/sec

Title: US-09-641-831-1
Perfect score: 573
Sequence: 1 atgatggagcaccatgaaga.....gagcctattgaagatga 573

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:
1: gb_da:
2: gb_hg:
3: gb_in:
4: gb_om:
5: gb_ov:
6: gb_pat:
7: gb_ph:
8: gb_pl:
9: gb_pr:
10: gb_ro:
11: gb_sts:
12: gb_sy:
13: gb_un:
14: gb_vl:
15: em_da:
16: em_fun:
17: em_hum:
18: em_in:
19: em_om:
20: em_or:
21: em_ov:
22: em_pat:
23: em_ph:
24: em_pl:
25: em_ro:
26: em_sts:
27: em_sy:
28: em_un:
29: em_vl:
30: em_hgo_hum:
31: em_hgo_inv:
32: em_hgo_rod:
33: em_hgo_hum:
34: em_hg_inv:
35: em_hg_rod:
36: em_hg_other:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	573	100.0	573	6	AX088023 Sequence
2	323	56.4	327	6	AX088025 Sequence
3	323	56.4	402	6	AX088027 Sequence
4	253	44.2	156375	2	AF002859 Homo sapi
5	253	44.2	159488	2	AC025141 Homo sapi
6	209	36.5	6423	10	AF199421 Mus muscu
7	170	29.7	163335	2	AC046189 Homo sapi
8	97.4	17.0	7445	9	HSAPX1
9	85.8	15.0	6014	9	AB033028 Homo sapi
10	64.6	11.3	196542	9	AC002365 Homo sapi
11	45	7.9	156413	9	HSAL19B20 Human DNA
12	44.8	7.8	39213	9	AC003047 Homo sapi
13	44.8	7.8	44400	9	AC090481 Homo sapi
14	43.6	7.6	173510	9	AL359272 Human DNA
15	37.8	6.6	2111	9	S39392 Homo sapien
16	37.8	6.6	3984	9	HOMCAP
17	37.4	6.5	1567	9	AF345905 Homo sapi
18	37.4	6.5	1669	9	AF345904 Homo sapi
19	37.4	6.5	1686	9	AF345906 Homo sapi
20	36.8	6.4	1485	10	AF002283 Mus muscu
21	36.8	6.4	329709	1	AP002997 Mesothelio
22	36.6	6.4	3220	9	BC008741 Homo sapi
23	36.6	6.4	3287	9	AF061258 Homo sapi
24	36.6	6.4	207079	2	AC015743 Homo sapi
25	36.6	6.4	209749	2	AC015564 Homo sapi
26	36.2	6.3	1696	10	AF095585 Rattus no
27	35.8	6.2	1724	9	AF265209 Homo sapi
28	35.8	6.2	1725	9	HUMENISMA
29	35.8	6.2	1757	9	BC001093 Homo sapi
30	35.8	6.2	1896	9	RNU48247
31	35.8	6.2	7218	6	I66494
32	35.8	6.2	134482	9	AC007335 Homo sapi
33	35.8	6.2	160460	2	AC020766 Homo sapi
34	35.6	6.2	112027	9	AC007006 Homo sapi
35	35.6	6.2	187727	2	AC064806 Homo sapi
36	35	6.1	2283	9	AB006955 Homo sapi
37	34.8	6.1	1282	5	SSA308531 Salmo sal
38	34.8	6.1	147197	2	AC091912 Homo sapi
39	34.8	6.1	171280	2	AC024186 Homo sapi
40	34.8	6.1	183349	2	AC027501 Homo sapi
41	34.8	6.1	237308	2	AC009593 Homo sapi
42	34.6	6.0	1191	10	RNRIT18
43	34.6	6.0	4542	3	AB036856 Homo sapi
44	34.4	6.0	1607	9	HSAL13766 Homo sapi
45	34.4	6.0	195630	2	AC087898 Mus muscu

ALIGNMENTS

RESULT 1

AX088023 573 bp DNA PAT 17-MAR-2001
LOCUS AX088023
DEFINITION Sequence 1 from Patent WO0114422.
ACCESSION AX088023
VERSION AX088023.1 GI:13396951
KEYWORDS

SOURCE

ORGANISM human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

1 (Bases 1 to 573)
AUTHORS Turner C.A., Zambrowicz B., Friedrich G., Nehls M. and Sands A.T.
TITLE Sequence derived from a human mammary gland cDNA library
JOURNAL Patent: WO 0114422-A 1 01-MAR-2001;
Lexicon Genetics Incorporated (US)

FEATURES

source
1..573
/organism="Homo sapiens"
/db_xref="taxon:9606"

```

BASE COUNT      151 a      152 c      152 g      118 t
ORIGIN

Query Match      100.0%; Score 573; DB 6; Length 573;
Best Local Similarity 100.0%; Pred. No. 5.5e-159;
Matches 573; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atgatgaggaccactgaagacttcacaaagcctagtgccacattaaacttaacacagcc 60
    |||
Db 1 ATGATGAGGACCCTGAAGACTTCCACAAAGCCTAGTGCCACATTAACCTTAACACGGCC 60
    |||

QY 61 accaagggaaggtacatttatctggagcattctctggagggagagctccctggggttt 120
    |||
Db 61 ACCAAGGGAAGGTACATTTATCTGGAGGCATTCCTGGAGGGAGGAGCTCCCTGGGGTTTT 120
    |||

QY 121 actctaaagggtggcctggagcagcggagaaaccattatcatctctaaaggtcgaagaagg 180
    |||
Db 121 ACTCTAAAGGGTGGCCTGGAGGCATTCCTGGAGGGAGGAGCTCCCTGGGGTTTT 120
    |||

QY 181 ggcgaagcagacacccctgagctccaaactcagcctgagctgggagtgaggttgacatcaat 240
    |||
Db 181 GGCAAGCAGACACCCCTGAGCTCCAACTCGAGGCTGGGATGAGGTGTGCACATCAAT 240
    |||

QY 241 gagtgactctgagcagctccagaaaggagcaggtttccctggtaagagatcctacaag 300
    |||
Db 241 GAGTGACTCTGAGCAGCTCCAGAAAGGAGGAGTTCCTGGTGAAGGATCCTACAAG 300
    |||

QY 301 accctcaggtgtagtgagcagcctctccaccggcagcctgttagcctcgagttgac 360
    |||
Db 301 ACCCTCAGGCTGAGTGCGCAGCCTCTCCACCAGGCTACTGTTAGCCTCGAGTTTGAC 360
    |||

QY 361 cctcaacatcccgagagctcctctagcagctcgaacctcatttagtctctactgct 420
    |||
Db 361 CCTCAACATCCCGAGAGATGCTCTCTAGGACTCGAACTCATTTAGTGCTCTACTGCT 420
    |||

QY 421 gatggagccatgagtgagctgtcaccacccctgggtgaagtggtggtctccacgctcc 480
    |||
Db 421 GATGGAGCCATGATGGAGCTGTGACACCTTGGGTGAAGTGGTGTCTCCACGTCCTC 480
    |||

QY 481 acctgggagcagcagtgccacagaaggtgtatctacccccaccagcagcacacatgc 540
    |||
Db 481 ACCTGGGAGCAGCAGTGCCGACAGAAGGTGTATCTACCCACCCAGCAGCACACATGC 540
    |||

QY 541 agaatctcaaaagagcctatttaagtagatga 573
    |||
Db 541 AGAATTTCAAAAGAGCCTATTAAAGTAGATGA 573
    |||

RESULT 2
AX088025 327 bp DNA PAT 17-MAR-2001
LOCUS
DEFINITION Sequence 3 from Patent WO0114422.
ACCESSION AX088025
VERSION AX088025.1 GI:13396952
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 327)
AUTHORS Turner,C.A., Zambrowicz,B., Friedrich,G., Nehls,M. and Sands,A.T.
TITLE Sequence derived from a human mammary gland cdna library
JOURNAL Patent: WO 0114422-A 3 01-MAR-2001;
Lexicon Genetics Incorporated (US)
FEATURES
    Location/Qualifiers
        source
            1..327
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
BASE COUNT      91 a      76 c      96 g      64 t
ORIGIN

```

```

Query Match      56.4%; Score 323; DB 6; Length 327;
Best Local Similarity 100.0%; Pred. No. 7.3e-85;
Matches 323; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atgatgaggaccactgaagacttcacaaagcctagtgccacattaaacttaacacagcc 60
    |||
Db 1 ATGATGAGGACCCTGAAGACTTCCACAAAGCCTAGTGCCACATTAACCTTAACACGGCC 60
    |||

QY 61 accaagggaaggtacatttatctggagcattctctggagggagagctccctggggttt 120
    |||
Db 61 ACCAAGGGAAGGTACATTTATCTGGAGGCATTCCTGGAGGGAGGAGCTCCCTGGGGTTTT 120
    |||

QY 121 actctaaagggtggcctggagcagcggagaaaccattatcatctctaaaggtcgaagaagg 180
    |||
Db 121 ACTCTAAAGGGTGGCCTGGAGGCATTCCTGGAGGGAGGAGCTCCCTAGGTCGAAGAGGG 180
    |||

QY 181 ggcgaagcagacacccctgagctccaaactcagcctgagctgggagtgaggttgacatcaat 240
    |||
Db 181 GGCAAGCAGACACCCCTGAGCTCCAACTCGAGGCTGGGATGAGGTGTGCACATCAAT 240
    |||

QY 241 gagtgactctgagcagctccagaaaggagcaggtttccctggtaagagatcctacaag 300
    |||
Db 241 GAGTGACTCTGAGCAGCTCCAGAAAGGAGGAGTTCCTGGTGAAGGATCCTACAAG 300
    |||

QY 301 accctcaggtgtagtgagcagcctctccaccggcagcctgttagcctcgagttgac 323
    |||
Db 301 ACCCTCAGGCTGAGTGCGCAG 323
    |||

RESULT 3
AX088027 402 bp DNA PAT 17-MAR-2001
LOCUS
DEFINITION Sequence 5 from Patent WO0114422.
ACCESSION AX088027
VERSION AX088027.1 GI:13396953
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 402)
AUTHORS Turner,C.A., Zambrowicz,B., Friedrich,G., Nehls,M. and Sands,A.T.
TITLE Sequence derived from a human mammary gland cdna library
JOURNAL Patent: WO 0114422-A 5 01-MAR-2001;
Lexicon Genetics Incorporated (US)
FEATURES
    Location/Qualifiers
        source
            1..402
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
BASE COUNT      110 a      91 c      117 g      84 t
ORIGIN

Query Match      56.4%; Score 323; DB 6; Length 402;
Best Local Similarity 100.0%; Pred. No. 7.4e-85;
Matches 323; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atgatgaggaccactgaagacttcacaaagcctagtgccacattaaacttaacacagcc 60
    |||
Db 1 ATGATGAGGACCCTGAAGACTTCCACAAAGCCTAGTGCCACATTAACCTTAACACGGCC 60
    |||

QY 61 accaagggaaggtacatttatctggagcattctctggagggagagctccctggggttt 120
    |||
Db 61 ACCAAGGGAAGGTACATTTATCTGGAGGCATTCCTGGAGGGAGGAGCTCCCTGGGGTTTT 120
    |||

QY 121 actctaaagggtggcctggagcagcggagaaaccattatcatctctaaaggtcgaagaagg 180
    |||
Db 121 ACTCTAAAGGGTGGCCTGGAGGCATTCCTGGAGGGAGGAGCTCCCTAGGTCGAAGAGGG 180
    |||

QY 181 ggcgaagcagacacccctgagctccaaactcagcctgagctgggagtgaggttgacatcaat 240
    |||
Db 181 GGCAAGCAGACACCCCTGAGCTCCAACTCGAGGCTGGGATGAGGTGTGCACATCAAT 240
    |||

QY 241 gagtgactctgagcagctccagaaaggagcaggtttccctggtaagagatcctacaag 300
    |||
Db 241 GAGTGACTCTGAGCAGCTCCAGAAAGGAGGAGTTCCTGGTGAAGGATCCTACAAG 300
    |||

QY 301 accctcaggtgtagtgagcagcctctccaccggcagcctgttagcctcgagttgac 323
    |||
Db 301 ACCCTCAGGCTGAGTGCGCAG 323
    |||

```

```

/chromosome="4"
/map="4q16-q20"
/clonetype="22SE15"

BASE COUNT      45683 a 32392 c 31715 g 45174 t 1411 others
ORIGIN

Query Match          44.2%; Score 253; DB 2; Length 156375;
Best Local Similarity 100.0%; Pred. No. 5,3e-64;
Matches    253; Conservative   0; Mismatches     0; Indels       0; Gaps         0;

OY  321 gagcttcctccaccggtacatcgtttagctcgagtttgaccttaaacatcccagaaggat 380
DB  31166 CAGCCTCTCCCAACCGGTACGCTTAGTCCTCGATTGTGAACCCTCAACAATCCCACAGAAGAT 31225
DB  31226 GCCTCTAGACTCGAAGCCTCATTTTGGTGCTCTACTGCTGTAAGACGCCATAAGTGAG 31285
OY  441 cgtgcgccactctgggtgaagtfgtgrgtctcaagtcctcacctggcagacgatgagcc 500
DB  31286 CTGTGCACCACTTGCGTAAAGTGTGTCTCCACGTCCCACTGGGACGACAGATGGCC 31345
OY  501 acagaagaattgatatcatccccccacagacaacacatgagaattccaagaagacta 560
DB  31346 ACGAAGAAGTTGATATCTACCCCCCACGACACACATGACAGAAATTCAAAAGACCTTA 31405
OY  561 tttaaataagatga 573
DB  31406 TTTAAGTAGATGA 31418

RESULT  5
LOCUS   AC025141
DEFINITION Homo sapiens clone RP11-224D4, WORKING DRAFT SEQUENCE, 8 unordered pieces.
ACCESSION AC025141
VERSION   AC025141.4 GI:10280891
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE   human.
ORGANISM Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 159488)
AUTHORS  Birren,B., Linton,L., Nussbaum,C. and Lander,E.
TITLE    Homo sapiens, clone RP11-224D4
JOURNAL  Unpublished
REFERENCE 2 (bases 1 to 159488)
AUTHORS  Birren,B., Linton,L., Nussbaum,C., Lander,E., Abraham,H., Allen,N.,
           Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
           Boguslavsky,I.L., Boukhalter,B., Brown,A., Burkett,G.,
           Campilano,A., Castelle,A., Chepel,Y., Colangelo,M., Collins,S.,
           Collamore,A., Cooke,P., Dearlano,K., Dewar,K., Diaz,J.S.,
           Dodge,S., Domino,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,
           Galagan,J., Gardyana,S., Glnde,S., Goyette,M., Graham,L.,
           Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
           Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
           Klein,J., Lahogue,K., Lamazares,R., Landers,T., Lehoczy,J.,
           Levine,R., Liu,C., Liu,G., Locke,K., MacDonald,P., Marquis,N.,
           McCarthy,M., McEwan,P., McGurt,A., McKernan,K., McNeeeters,R.,
           Meldrum,J., Menous,L., Mihova,T., Miranda,C., Mlenka,V., Morrow,J.,
           Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
           O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
           Pisanic,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
           Roy,A., Santos,R., Schaner,S., Severy,P., Spencer,B.,
           Stancio-Tromann,N., Stanojovic,N., Subramanian,A., Talamas,J.,
           Teffaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
           Vassilev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.-J.,
           Young,G., Zainoun,J., Zimmer,A. and Zody,M.
```

COMMENT

On Sep 23, 2000 this sequence version replaced gi:8077070.
All repeats were identified using RepeatMasker.
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu
----- Project Information
Center project name: L7938
Center clone name: 224_D_4

----- Summary Statistics
Sequencing vector: M13; M7815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 155108 bases at least Q40
Consensus quality: 157181 bases at least Q30
Consensus quality: 158099 bases at least Q20
Insert size: 150000; agarose-fp
Insert size: 158788; sum-of-contigs
Quality coverage: 6.8 in Q20 bases; agarose-fp
Quality coverage: 6.4 in Q20 bases; sum-of-contigs

* NOTE: this is a 'working draft' sequence. It currently
* consists of 8 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 11564: contig of 11564 bp in length
* 11565 11664: gap of 100 bp
* 11665 14943: contig of 3279 bp in length
* 14944 15043: gap of 100 bp
* 15044 19438: contig of 4395 bp in length
* 19439 19538: gap of 100 bp
* 19539 31843: contig of 12305 bp in length
* 31844 31943: gap of 100 bp
* 31944 98274: contig of 66331 bp in length
* 98275 98374: gap of 100 bp
* 98375 114530: contig of 16176 bp in length
* 114531 114650: gap of 100 bp
* 114651 147429: contig of 32779 bp in length
* 147430 147529: gap of 100 bp
* 147530 159488: contig of 11959 bp in length.

FEATURES
Source Location/Qualifiers

1..159488
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="RP11-224D4"
/clone_lib="RP11-224D4"
1..11564
/note="assembly_fragment"

misc_feature
vector_side:left
11665..114943
/note="assembly_fragment"

misc_feature
15044..19438
/note="assembly_fragment"

misc_feature
19539..31843
/note="assembly_fragment"

misc_feature
31944..98274
/note="assembly_fragment"

misc_feature
98375..114530
/note="assembly_fragment"

misc_feature
114651..147429
/note="assembly_fragment"

misc_feature
147530..159488
/note="assembly_fragment"

BASE COUNT 46031 a 33033 c 33024 g 46700 t 700 others
ORIGIN

Query Match 44.2% Score 253; DB 2; Length 159488;
Best Local Similarity 100.0%; Pred. No. 5.3e-64;
Matches 253; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 321 cagcctctccccaccggtcactgttagctgagttgacctcaacatccccagagat 380
|||||
Db 2903 CAGCCTCTCCCCACCGGTCACCTGTTAGCTCGAGTTGACCTCAACATCCCAGAGAT 2962
Qy 381 gcctctaggactggaacctcatttagtctctactgctgtagcgcctagtgag 440
|||||
Db 2963 GCCCTCTAGGACTCGAACCTCATTTAGTGTCTCTACTGTGATGGACGCATGAGTGGAG 3022
Qy 441 ctgtcgaccacacctgggtgaagtgtgtctccagctccacacctgggagacagatgccc 500
|||||
Db 3023 CTGTGACCACCTTGGGTGAAGTGTGTCTCCAGTCCCACTGGGACGACGATGGCC 3082
Qy 501 acagaaaggttgtatctaccacccacccagcacacacatgcagaaatttcaaaagagccta 560
|||||
Db 3083 ACAGAAAGGTTGTATCTACCCACCACGACACACATGCAGAAATTTCAAAGAGCCTA 3142
Qy 561 tttaagtagatga 573
|||||
Db 3143 TTTAAGTAGATGA 3155

RESULT 6

LOCUS AF199421 6423 bp mRNA ROD 13-DEC-1999
DEFINITION Mus.musculus PDZ domain actin binding protein Shroom mRNA, complete cds.
ACCESSION AF199421
VERSION AF199421.1 GI:6467989
KEYWORDS house mouse.
SOURCE Mus musculus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 6423)
AUTHORS Hilbrand, J.D. and Soriano, P.
TITLE Shroom, a PDZ domain-containing actin-binding protein, is required for neural tube morphogenesis in mice
JOURNAL Cell 99 (5), 485-497 (1999)
MEDLINE 20055594
PUBMED 10589677
REFERENCE 2 (bases 1 to 6423)
AUTHORS Hilbrand, J.D. and Soriano, P.
TITLE Direct Submission
JOURNAL Submitted (27-OCT-1999) Basic Sciences, Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North, Seattle, WA 98109-1024, USA

FEATURES
source Location/Qualifiers
1..6423
/organism="Mus musculus"
/db_xref="taxon:10090"
233..6193
/note="shrm"
/codon_start=1
/product="PDZ domain actin binding protein Shroom"
/protein_id="AAF13269.1"
/db_xref="GI:6467990"
/translation="MKTPENLEPSPATNPSTPTTFVYLEALLEGAPWGTLLKGG
LERGEPLIISKIEGKADSVSGLQAGDEVTHINEVALSSPRRAVSILVKGSYKTLR
LVVRDVCAAPGHADPGTSGKSLSELLTCSPOHRKATWSGVKRLKORCSPPAPRPH
SWHTTKFGETQPDVSMOISOGTGMPPHQSHTSSSSSDLSNHYALRRSPDCSS
QGSMELEPSSGYPCHLLSPAKSTSSIDQLGHLNKRDSAYSSSTSSSIFPPPG
GSARSGSMDVISARGLLGEMQRADIRVYTVYDRGVSSVEYVAPSAALLQGRD
AHASRSGCAKHSIPKGGTTPSPSQOCSGLSTATDNLPGAGAPLPPTRSDSY
AAFRHRERPSWSLDRFCRPPQNSGSKTPFAEDQLHTVPRSPSPNSPPVSKKH
NYTKAQPGQPLLPTGIYVPSPPEPFAQVPQSPSSNGTVYFALVKESGYTAAGTC

Page 7

RESULT	9
AB033028	AB033028 6014 bp mRNA PRI 11-NOV-1999
LOCUS	
DEFINITION	Homo sapiens mRNA for KIAA1202 protein, partial cds.
ACCESSION	AB033028
VERSION	AB033028.1 GI:6330421
KEYWORDS	
SOURCE	Homo sapiens brain cDNA to mRNA, clone_11b:pbunescript11 SK plus clone:fg03353.
ORGANISM	Homo sapiens
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE	1 (sites)
AUTHORS	Nagase,T., Ishikawa,K., Kikuno,R., Hiroseawa,M., Nomura,N. and Ohara,O.

QY	83	tgagagcaattcttggaaaggagagctccctggggtttactctaaagggttggccggagc	142
Db	170	tggagctgacgttgagcggcggccggccctggggccttcacctgaaaggcggccggagc	229
QY	143	acggagaaccatctaactctcttaagctcgaaaggaggccaagcagacacctgagct	202
Db	230	acggcgagccctcgctgctatccacacagattgaagaggcagctaaagc---cggcgggcgctg	286
QY	203	ccaaacctgcaagcttgggagatgaggttgtgcacataatgaggtgacctctgagcagctcca	262
Db	287	acaaacttactggctggaatgagatgcctgcgcattcaatgacattgctctcagagggttta	346
QY	263	gaagagagcagattccctctgtgtaagaagtaacctaaagaacctctgagctgtagt	317
Db	347	gacagagaagcagtttgctctggtgaaagggtctccattaaagacctgaaagctggctgct	401

FEATURES	Location/Qualifiers
source	1. .6014

2 (bases 1 to 6014)
Ohara O., Nagase T. and Kikuno R.
Direct Submission
Submitted (04-OCT-1999) to the DDBJ/EMBL/GenBank databases. Osamu
Ohara, Kazusa DNA Research Institute, Laboratory of DNA Technology
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
(E-mail: cdnainfo@kazusa.or.jp, URL: <http://www.kazusa.or.jp/huge/>
Tel: +81-438-52-3913, Fax: +81-438-52-3914)
/01/01/1999

gene	15. .4523
CDS	/gene="KIAA1202" <15. .4523

BASE COUNT	1613 a	1620 c	1535 g	1246 t
ORIGIN				

1

```

/Note="Similar to sequence with GenBank Accession number
236840"
repeat_region      15384..15565
                    /rpt_family="MER58B"
repeat_region      15656..15956
                    /rpt_family="AluSg"
repeat_region      16091..16660
                    /rpt_family="L1MB7"
                    /join(<17023..117176,34350..34481,37731..40071,41563..41663,
75548..76243,80507..81058,82568..82739,88014..88286,
90044..92814)
                    /gene="HSAAPL"
                    /product="APPL"
                    /join(<17023..117176,34350..34481,37731..40071,41563..41663,
75548..76243,80507..81058,82568..82739,88014..88286,
90044..92814)
                    /gene="HSAAPL"
                    /protein_id="AAC32592.1"
                    /db_xref="gi:1773381"
                    /translation="TEGSKAAVADKILAGDEIVGINDIGSGFROEATCLVKGSHKT
LKVYKRSLGMRPHSHATKESDSHELAASPTSTSGCPMSGRHASSHSDLS
SSMOQNTQRTLDHSSLSGVSVDHPSRLSVAKSNSIDHLSHSHKSCPPRRV
SSSPDHLSKADTSSAENILYTVGLWAPRQGGQAQACDPQSEKLCSPRRV
GDSGPRPEYNAEKLAPGRSNFVWYVDKKAPSSPPPPPLRSQFAATKS
HEKAGDPVSEAAAHOFTALAOAPRGDRPELDWRASHPGLSGGSGPCQP
EAAHDSNPPSKDASRLQASISDVRFPOSPHSGHPLYSHPSLCADLQEP
GAASFQNDSPQVRLSSCDOKLGSOGPPRCVGOAOALMGACMPDIALALE
SLPPTVQSGSPRHLPQEPGPPDARETCRPLDGAGCGCASAQOEPRAAEKAS
RLAASITWADGSSRICPEETPLHSLTQEGRRPSSPESATPPPPDPAVGPTR
KSDPATLRLNEIOMHRAKIQRSSTVALTAGEDGTGRRAOLGGTQGLAPGT
YKDHLEQAQVRLATSPFRKRDIDPDGLTSELEHMGDPDVPQHRWEALQAP
SISGSPHPRIGGRRTAEOKIKSSEPKANEVGLRGYSPHOPRTSEDTGSLN
DMKFEETSKYVPPRPAKQKIQSKQKRIARASGRCHSADLIDVLDPOER
POHVGKRSNSPSTHYKQEAVELRKQGDGEPEBELPSVRAEBCSTROADQ
CREGSGSQHPQKAPNPPTSELSHCRGAPELPRREGSGAGTLPRYKTSSEETP
ADIGRAPOGSPGLHARGODSWPVSSALSKRPAPQRPPTKATCERGSQHVSDA
SPGVIGRPPTSPASIDVYVARLSHSBVSFSSAQODPTKATCERGSQHVSDA
SRPLPALLPQQHRLQOTAMETSRSPQAPQKLDKPKLIDODSDRIEFVM
DNNTYKMPVPIKIVHESQPEKESROSLACPAPEKRTSPGSLYMAKAKTYDIAKSE
PLTYRQGAPEPAPRAQAPPEPQSTQVPEKDRCTSPGSLYMAKAKTYDIAKSE
ELAREIVGDKSLADIDPSVKIKTMDMEGIFPDHLEHLEPQHRKLLKPTBPB
SPEERKEVSPAAVSLATNSTRYSTSAKAEELLIMKDLQEOHEHESGSLDLDL
SVYKQRLTISIKQLVLRARESLIEDVOANTVGAEEAIVKQCKSEFDEKME
IGDLKVVNLLTSLSRILARVAMLNLDGASPDROSLEKORVILQOHEAKELK
ENDRRERIVFDILANTLSESLADYEHYKMSALITIRELEDIHIGEEQLKCLL
DSLQPERGR"
repeat_region      complement(17976..18074)
                    /rpt_family="Hy3"
repeat_region      18162..18473
                    /rpt_family="AluJb"
repeat_region      19636..19717
                    /rpt_family="MER21A"
repeat_region      complement(19798..19937)
                    /rpt_family="L1MB3"
repeat_region      complement(20181..20327)
                    /rpt_family="FRAM"
repeat_region      complement(22367..22401)
                    /rpt_family="AT_rich"
repeat_region      22406..22461
                    /rpt_family="GGAAn"
                    complement(22462..22627)
                    /rpt_family="AluJo"

```

```

repeat_region      complement(22638..22933)
                    /rpt_family="AluSx"
repeat_region      complement(26881..26957)
                    /rpt_family="MER5A"
repeat_region      complement(2726..27762)
                    /rpt_family="POLY_A"
repeat_region      complement(27876..27977)
                    /rpt_family="AluJo/FRAM"
repeat_region      complement(28048..28069)
                    /rpt_family="AT_rich"

Query Match      11.38; Score 64.6; DB 9; Length 196542;
Best Local Similarity 65.7%; Pred. No. 3.9e-08;
Matches 94; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

OY 175 gaaggggcaagaagcaccctagctcccaactgcagctgggagatgagttgac 234
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 17028 GAAGAGGCGACTAAGCCGCGCGGTGACAGATTACTGGCTGAGATGAGATCGCGC 17087

OY 235 atcaatgagtgactctgagcagctccgaagaaggagcagttccctgggaaagatcc 294
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 17088 ATCAATGACATTGCTCTCAAGGTTTACAGACAGAGCATTTGCTGTGTAAGGGGTC 17147

OY 295 tacaagacctcagctgtagt 317
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 17148 CATAGACCTCGAAGCTGCTCT 17170

```

```

RESULT 11
LOCUS      HSA119E20
DEFINITION Human DNA sequence from clone RP11-119E20 on chromosome
Xp11.21-11.23 Contains part of the gene for KIAA1202 protein, ESTs,
STSs and GSSs, complete sequence.
ACCESSION  AL121865
VERSION     AL121865.7 GI:8979790
KEYWORDS    HTG: KIAA1202.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE   1 (bases 1 to 156413)
AUTHORS     Wilson,S.
TITLE       Direct Submission
JOURNAL     Submitted (27-JUL-2000) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
            requests: clonerequest@sanger.ac.uk
COMMENT     On Jul 8, 2000 this sequence version replaced gi:8977862.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence has been finished according to sequence map criteria
            as follows. An attempt is made to resolve all sequencing problems,
            such as compressions and repeats, but not necessarily within known
            annotated human repeat sequence elements (e.g. Alu). Where the
            sequence is ambiguous, there is an annotation using the 'unsure'
            feature key.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em: EMBL; SW: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C.elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome X, constructed by the Sanger Centre Chromosome X Mapping
            Group. Further information can be found at
            http://www.sanger.ac.uk/NGP/chrX
            RP11-119E20 is from the library RP11-11.1 constructed at the
            Roswell Park Cancer Institute by the group of Pieter de Jong. For
            further details see http://bacpac.med.buffalo.edu/
            VECTOR: pBACe3.6

```

FEATURES

This sequence is the entire insert of clone RP11-119E20.
Location/Qualifiers

FEATURES
source

[illegible]


```

misc_feature      11146..11196
note="low coverage"
repeat_region    11203..11335
/rpt_family="FRAM"
repeat_region    11351..11652
/rpt_family="AluSx"
complement(11351..11652)
repeat_region    12228..12254
/rpt_family="AT-rich"
complement(12228..12254)
repeat_region    12257..12556
/rpt_family="AluSx"
complement(12257..12556)
repeat_region    12557..12590
/rpt_family="AluSx"
complement(12557..12590)
repeat_region    12590..12994
/rpt_family="AT-rich"
complement(12590..12994)
repeat_region    12994..13020
/rpt_family="AluJo"
complement(12994..13020)
repeat_region    13023..13093
/rpt_family="(GA)n"
complement(13023..13093)
repeat_region    13667..13854
/rpt_family="MER58A"
complement(13667..13854)
repeat_region    14127..14233
/rpt_family="AluSx"
complement(14127..14233)
repeat_region    14368..14670
/rpt_family="MER5A"
complement(14368..14670)
repeat_region    15204..15494
/rpt_family="AluSx"
complement(15204..15494)
repeat_region    16134..16246
/rpt_family="AluJb"
complement(16134..16246)
repeat_region    16250..16538
/rpt_family="MER4_internal"
complement(16250..16538)
repeat_region    16660..16775
/rpt_family="AluJb"
complement(16660..16775)
repeat_region    17520..17805
/rpt_family="AluJb"
complement(17520..17805)
repeat_region    18199..18272
/rpt_family="AluJb"
complement(18199..18272)
repeat_region    19058..19350
/rpt_family="MIR"
complement(19058..19350)
repeat_region    19372..19584
/rpt_family="AluJb"
complement(19372..19584)
repeat_region    19599..19778
/rpt_family="AluJb"
complement(19599..19778)
repeat_region    19778..19788
/rpt_family="MER58A"
complement(19778..19788)
repeat_region    19788..19796
/rpt_family="MER58A"
complement(19788..19796)
repeat_region    19796..19800
/rpt_family="MER58A"
complement(19796..19800)
repeat_region    20211..20264
/rpt_family="AluJb"
complement(20211..20264)
repeat_region    20280..20314
/rpt_family="POLY_A"
complement(20280..20314)
repeat_region    20415..20468
/rpt_family="FAM"
complement(20415..20468)
repeat_region    20472..20771
/rpt_family="AluJb"
complement(20472..20771)
repeat_region    21254..21353
/rpt_family="AluJb"
complement(21254..21353)
repeat_region    21386..21447
/rpt_family="MT1D"
complement(21386..21447)
repeat_region    22090..22391
/rpt_family="AluSx"
complement(22090..22391)
repeat_region    22393..22601
/rpt_family="L2"
complement(22393..22601)
repeat_region    23070..23235
/rpt_family="AluJb"
complement(23070..23235)
repeat_region    23484..23618
/rpt_family="AluJb"
complement(23484..23618)
repeat_region    23620..23920
/rpt_family="AluJb"
complement(23620..23920)
repeat_region    23921..24091
/rpt_family="AluJb"
complement(23921..24091)
repeat_region    24131..24432
/rpt_family="AluJb"
complement(24131..24432)
repeat_region    24551..24787
/rpt_family="AluJb"
complement(24551..24787)

```

```

/rpt_family="AluJo"
25233..25550
/rpt_family="AluSp"
25566..25866
/rpt_family="AluJb"
25867..25901
/rpt_family="(GAAA)n"
complement(25867..25901)
/rpt_family="LPR12"
complement(25867..25901)
/rpt_family="LMD2"
28409..28490

```

```

Query Match      7.8%; Score 44.8; DB 9; Length 39213;
Best Local Similarity 69.3%; Pred. No. 0.028; 27; Indels 0; Gaps 0;
Matches 61; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

QY 83 tggagcattcttgcggaggagcgtccctgggttttacttaagggtgacctgagc 142
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 29327 TGGAGTGCAGTGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 142
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 143 acggagaaccatttaatactctcttaagggt 170
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 29387 ACGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 170
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 13
AC090481 44400 bp DNA PRI 01-APR-2001
LOCUS Homo sapiens Xp BAC RP11-8H7 (Roswell Park Cancer Institute Human
AC090481 BAC library) complete sequence.
AC090481 AC090481.9 GI:13491204
VERSION HTG.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Muzny,D.M., Adams,C., Bailey,M., Barberia,J., Blankenburg,K.,
Bodota,B., Bouck,J., Bowie,S., Brooks,A., Brundage,B., Buhay,C.,
Bunac,C., Burkett,C., Burrows,J., Carter,M., Chacko,J., Chen,Z.,
Cox,C., David,R., Delgado,O., Deshazo,D., Di,W., Ding,Y.,
Domah-Rashid,N., Dugan-Rocha,S., Durbin,K.J., Fernandez,C.,
Ferraguto,D., Forcum-Tansey,J., Frantz,P., Ganesh,R., Garcia,C.,
Gorrell,J.H., Gorrell,L.L., Guevara,W., Harris,K., Jackson,L.E.,
Hodgson,A., Hogue,M., Hollway,C., Hosak,H., Jackson,L.E.,
Jackson,L., Jones,M., Kelly,S., Kondejewski,N., Kong,Y., Kovar,C.,
Leal,B., Li,Z., Lichtarge,O., Liu,J., Liu,W., Logan,O.,
Lozano,R.J., Lu,J., Lucier,R., Martin,R., Martinez,C., McLeod,M.P.,
Mei,G., Morgan,M., Morris,S., Nash,S., Nelson,A., Nguyen,R.,
Nguyen,N., Nguyen,S., Osval,G., Rampell,L.R., Parish,B., Paxton,S.,
Payton,B., Perez,L., Pu,L.L., Quiles,M., Rashid,N.D., Reiter,D.,
Rives,M., Samuel,S., Say,J., Scherer,S., Shah,E., Shen,H.,
Simon,M., Sparks,A., Stamps,A., Sugang,R., Tabor,P., Taylor,T.,
Vasquez,L., Vinson,R., Vo,Q., Wabnah,M., Watlington,S., Taylor,T.,
Weinstock,G., Weinstock,I.R., Williamson,A., Worley,K., Wren,J.,
Wrensford,G., Yu,W., Zhou,X., Nelson,D. and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 44400)
Worley,K.C.
Direct Submission
Submitted (24-FEB-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 44400)
Worley,K.C.
Direct Submission
Submitted (31-MAR-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
4 (bases 1 to 44400)

```


<pre>repeat_region repeat_region repeat_region STS</pre>	<pre>/rpt_family="L1MC5" complement(9676..9885) /rpt_family="MER58B" 9886..10166 /rpt_family="Aluwb" complement(10167..10258) /rpt_family="MER58B" 10294..10455 /standard_name="OAL-CA"</pre>	<p>Query Match Best Local Similarity 69.3%; Pred No. 0.026; Length 44400; Matches 61; Conservative 0; Mismatches 27; Indels 0; Gaps 0;</p>	<pre>OY 83 tggaggcattctgagggagagtccttgcgtttacttaaaagtgcctggagc 142 Db 32491 TGGAGGTGACGTGACGCCGGCGCCCTGCCTTACCCTGAAGCGGCCCGCAGC 32550 OY 143 acggagaacattcaatcatctctaaggt 170 Db 32551 ACGCGAGCGCGTGTCATCACCAAGGT 32578 </pre>	<pre>RESULT 14 AL359272/c LOCUS DEFINITION AL359272 ACCESSION VERSION KEYWORDS SOURCE ORGANISM</pre>	<pre>Human DNA sequence from clone RP11-554P16 on chromosome X. Contains the first coding exon of the gene KIAA1202, STSS, GSsS and a CpG island, complete sequence. AL359272 AL359272.9 GI:9650579 HTG; Cpg island; KIAA1202. human Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 173510) Direct Submission Submitted (28-SEP-2000) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk requests: clonerequest@sanger.ac.uk On Aug 1, 2000 this sequence version replaced gi:9588592. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above. This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous repeat sequence elements (e.g. Alu). Where the feature key. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Embl, EMBL; Swi, SWISSPROT; Trl, TREMBL; Wp, WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C-elegans/wormpep This sequence was generated from part of bacterial clone contigs of human Chromosome X, constructed by the Sanger Centre Chromosome X Mapping Group. Further information can be found at http://www.sanger.ac.uk/HGP/ChrX RP11-554P16 is from the library RP11-11.2 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see http://bacpac.med.buffalo.edu/ VECTOR: pBAC3.6 This sequence is the entire insert of clone RP11-554P16. Location/Qualifiers 1..173510 /organism="Homo sapiens" /db_xref="taxon:9606"</pre>	<pre>repeated_region misc_feature repeated_region repeated_region repeated_region repeated_region repeated_region repeated_region repeated_region repeated_region repeated_region repeated_region repeated_region repeated_region repeated_region repeated_region misc_feature misc_feature repeated_region misc_feature misc_feature repeated_region</pre>	<pre>/chromosome="X" /clone="RP11-554P16" /clone_lib="RPCI-11.2" 2..192 /note="L1MEC repeat: matches 1913..2098 of consensus" 7..674 /note="match: GSS: Em:AQ373815" 195..511 /note="AlusP repeat: matches 1..312 of consensus" 543..845 /note="Alusx repeat: matches 1..304 of consensus" 950..1127 /note="L1MEC repeat: matches 1968..2136 of consensus" 1190..2306 /note="L1M4b repeat: matches -180..888 of consensus" 2425..3343 /note="L1M4 repeat: matches 3078..4092 of consensus" 3344..3717 /note="MSTD repeat: matches 1..394 of consensus" 3718..4412 /note="L1M4 repeat: matches 4092..4274 of consensus" 4413..4452 /note="THELB repeat: matches 1..42 of consensus" 4453..4763 /note="Alusx repeat: matches 1..310 of consensus" 4764..5111 /note="THELB repeat: matches 42..364 of consensus" 5112..5774 /note="L1M4 repeat: matches 4274..5344 of consensus" 5775..6080 /note="AlusG repeat: matches 1..307 of consensus" 6081..6296 /note="L1M4 repeat: matches 5344..5581 of consensus" 6299..6923 /note="L1PAL10 repeat: matches 5525..6165 of consensus" 7394..7652 /note="L1M4 repeat: matches 2120..2383 of consensus" 7652..10378 /note="L1PA4 repeat: matches 3414..6142 of consensus" 10369..10878 /note="L1M4 repeat: matches 2350..2879 of consensus" 10875..11250 /note="L1 repeat: matches 4741..5131 of consensus" 11247..11619 /note="L1MA7 repeat: matches 5903..6284 of consensus" 11642..11764 /note="L1M4 repeat: matches 5351..5475 of consensus" 11764..11943 /note="L1 repeat: matches 3846..4036 of consensus" 12269..13359 /note="L1MC3 repeat: matches 6659..7739 of consensus" complement(13126..13490) /note="match: STS: Em:HS3082G9" 13430..13497 /note="L1M4 repeat: matches 1995..2062 of consensus" 13498..13803 /note="AluJo repeat: matches 3..307 of consensus" 13804..15085 /note="L1M4 repeat: matches 720..1995 of consensus" 15191..15376 /note="L1M4 repeat: matches 408..602 of consensus" 15565..16035 /note="L1M4 repeat: matches -252..225 of consensus" 16058..16114 /note="L2 repeat: matches 2686..2745 of consensus" 16128..16369 /note="L1MAL10 repeat: matches 5928..6318 of consensus" 17285..17533 /note="match: GSS: Em:AQ134667" 17455..17899 /note="match: GSS: Em:AQ142071" 17875..18018 /note="MIR repeat: matches 96..252 of consensus"</pre>
--	---	--	---	--	--	---	--

```

repeat_region 18979..19286
/note="AluJb repeat: matches 2..308 of consensus"
repeat_region 19391..19823
/note="L2 repeat: matches 2284..2708 of consensus"
repeat_region 20320..20610
/note="AluJx repeat: matches 1..289 of consensus"
misc_feature 21916..22121
/note="match: STS: Em:G03829"
repeat_region 23553..23862
/note="MER3 repeat: matches 1..307 of consensus"
repeat_region 23946..24118
/note="MER repeat: matches 77..259 of consensus"
repeat_region 24356..24475
/note="L2 repeat: matches 2624..2750 of consensus"
repeat_region 24929..25125
/note="MIR repeat: matches 23..256 of consensus"
repeat_region 25251..25386
/note="MER63 repeat: matches 2..768 of consensus"
repeat_region 25604..25656
/note="L2 repeat: matches 2657..2709 of consensus"
repeat_region 26047..27589
/note="L1MB8 repeat: matches 4591..6159 of consensus"
repeat_region 27530..27886
/note="AluJx repeat: matches 1..297 of consensus"
repeat_region 27887..28214
/note="L1MB8 repeat: matches 4271..4591 of consensus"
repeat_region 28258..29096
/note="L1MB8 repeat: matches 3264..4157 of consensus"
repeat_region 29118..29521
/note="MIR repeat: matches 1..383 of consensus"
repeat_region 29523..30381
/note="L1MB8 repeat: matches 2496..3231 of consensus"
repeat_region 30382..30685
/note="AluJb repeat: matches 2..239 of consensus"
repeat_region 30686..31069
/note="L1MB8 repeat: matches 2110..2496 of consensus"
repeat_region 31088..31421
/note="L1PA13 repeat: matches 5820..6156 of consensus"
repeat_region 31422..31617
/note="L1PA4 repeat: matches 5374..5552 of consensus"
repeat_region 31618..31909
/note="AluJx repeat: matches 1..294 of consensus"
repeat_region 31910..31976
/note="L1PA4 repeat: matches 5304..5374 of consensus"
repeat_region 31978..32359
/note="L1MB8 repeat: matches 1951..2330 of consensus"
repeat_region 32354..33592
/note="L1MAC repeat: matches 969..2197 of consensus"
repeat_region 33935..34790
/note="L1MB4 repeat: matches -266..602 of consensus"
repeat_region 34856..35219
/note="L1MB4 repeat: matches 1..364 of consensus"
repeat_region 36043..36159
/note="L2 repeat: matches 2576..2710 of consensus"
misc_feature 38455..39057
/note="match: GSS: Em:AQ008362"
misc_feature 41978..42373
/note="match: STS: Em:L18363"
repeat_region 42056..42087
/note="L6 copies 2 mer tg 96% conserved"
repeat_region 42311..42540
/note="MIR repeat: matches 19..262 of consensus"
repeat_region 42824..42914
/note="L2 repeat: matches 2606..2709 of consensus"
repeat_region 43119..43173
/note="L2 repeat: matches 2687..2742 of consensus"
repeat_region 43122..43185
/note="MIR repeat: matches 194..257 of consensus"
repeat_region 44700..44743
/note="L2 copies 2 mer aa 75% conserved"
repeat_region 45956..46137
/note="MIR repeat: matches 20..199 of consensus"
repeat_region 46401..46617

```

```

repeat_region 46729..47035
/note="MIR repeat: matches 23..262 of consensus"
repeat_region 47439..47682
/note="L1MB1 repeat: matches 5666..5914 of consensus"
repeat_region 48430..48527

Query Match 7.6%; Score 43.6; DB 9; Length 173510;
Best Local Similarity 67.8%; Pred. No. 0.066;
Matches 61; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

OY 94 ctggaaggagagctccctgggtttactcctaagggtgcttgagcagcaggaacca 153
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 84025 CTCGACAGGGGGCCACCTGGGCTTAAAGGGGGTCTGAAACACTGAGCCG 83966

OY 154 ttaatcatcttaaggttcgaagaaggagc 183
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 83965 CTCACAGTGTCTAAGTAAGAACTGGCGGC 83936

RESULT 15
LOCUS S39392 2111 bp mRNA PRI 05-MAR-2001
DEFINITION Homo sapiens protein tyrosine phosphatase mRNA, partial cds.
ACCESSION S39392
VERSION S39392.1 GI:250890
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 2111)
AUTHORS Arimura,Y., Hinoda,Y., Itoh,F., Takekawa,M., Tsujisaki,M.,
Adachi,M., Imai,K. and Yachi,A.
TITLE CDNA cloning of new protein tyrosine phosphatases in the human
colon
JOURNAL Tumour Biol. 13 (3), 180-186 (1992)
MEDLINE 92327504
GENBANK staff at the National Library of Medicine created this
entry [NCBI gi250890] from the original journal article.
REMARK This sequence comes from Fig. 3.
FEATURES
source Location/Qualifiers
1..2111
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CL6"
/rissue="Colon"
<1..>2111
/note="Prpase; This sequence comes from Fig. 3"
/codon_start=3
/product="protein tyrosine phosphatase"
/protein_id="AAB22439.2"
/db_xref="GI:13195754"
/translation="OHSGKQSEAESECYINARTDFYVELHSGRDLHNLDMTIGIA
SAGVAVYRKTCISPYPMYNITIKISRRKFFIHORQAESEHIVAFNMILNYSCK
NLKSCVEHHTFQAKLLPOEKNTVSOVWTKSVNNQYKAGVGMVNPDA
MRSISVEHLEKSLPSRSPPTPMRSRLRHRIRKPHSSADNANMTITTED
VFETVYKGLAPQSDSEVSQNRSPHQSLENNPASYSITQSSSSVSSNAPGSCS
PDGVDOOLDDFHRVTKGSTDASQYCDKNDNGSYLVLRITPDEDKGFNKG
GVQOKMPLVIRRAVRNPSPADTCIPKLNESDQYLINGRDISHTHQQVMFKKSES
HSRELAIVIRRAVRNPSPADTCIPKLNESDQYLINGRDISHTHQQVMFKKSES
GTVLIDPEQLYRKRPGLATITFATLPONLDKNRKYDLPTDTRVLLQGNEDYINASY
NMEIPANLVNKYIATQGLPHTCAQFMQVMDQKSLVMTLTITLTERGRTCHOYWP
DPPDVNMGHGFHIOCOSEDCITIVYSREMLVNTQGEETVTHLOIYAMPDNGIFDD
SDPLEFVNVRSLRVSEPLVHCSAGIGRTGVLVMTETACLTERRNLPTPLDLYR
KMDORAMVQTSQYKFC"
BASE COUNT 594 a 528 c 510 g 479 t
ORIGIN
Query Match 6.6%; Score 37.8; DB 9; Length 2111;
Best Local Similarity 53.8%; Pred. No. 2.9;
Matches 78; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

```

QY 97 gagggagagctccctggggttttactcctaaagggtggcctggagcacgagagaccattta 156
 Db 969 GATCAAGATGAAATTTGGATTATCTTAAGGGAGGAGTGGATCAAAAAGATGCCTCTT 1028
 QY 157 atcatctcttaaggtcgagaggggcaagagagacacacctgagctccaaactgcaggct 216
 Db 1029 GTGCTATCAAGGATTAACCCAGAGTCACCTGCGGACACCTTCTTAAGCTGAACGAA 1088
 QY 217 ggggatgaggtgtgtgcacatcaatg 241
 Db 1089 GGGATCAAAATCGTTAATCAATG 1113

Search completed: January 31, 2002, 19:45:37
 Job time: 1755 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd

OM nucleic - nucleic search, using SW model

```
Run on:      January 31, 2002, 19:50:27 ; Search time 163.78 Seconds
              (without alignments)
              2999.434 Million cell updates/sec
```

Title: US-09-641-831-1
 Perfect score: 573
 Sequence: 1 atgatgaggaacctgaaga.....gagcctatctaagtagatga 573

Searched: 930621 segs, 428662619 residues
Total number of hits satisfying chosen parameters: 1861242

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

```
Post-processing: Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries
```

Database :

N_GeneSeq_1101:*

1: /SID52/gcgdata/geneSeq/geneSeq/NA1980.DAT.*

2: /SID52/gcgdata/geneSeq/geneSeq/NA1981.DAT.*

3: /SID52/gcgdata/geneSeq/geneSeq/NA1982.DAT.*

4: /SID52/gcgdata/geneSeq/geneSeq/NA1983.DAT.*

5: /SID52/gcgdata/geneSeq/geneSeq/NA1984.DAT.*

6: /SID52/gcgdata/geneSeq/geneSeq/NA1985.DAT.*

7: /SID52/gcgdata/geneSeq/geneSeq/NA1986.DAT.*

8: /SID52/gcgdata/geneSeq/geneSeq/NA1987.DAT.*

9: /SID52/gcgdata/geneSeq/geneSeq/NA1988.DAT.*

10: /SID52/gcgdata/geneSeq/geneSeq/NA1989.DAT.*

11: /SID52/gcgdata/geneSeq/geneSeq/NA1990.DAT.*

12: /SID52/gcgdata/geneSeq/geneSeq/NA1991.DAT.*

13: /SID52/gcgdata/geneSeq/geneSeq/NA1992.DAT.*

14: /SID52/gcgdata/geneSeq/geneSeq/NA1993.DAT.*

15: /SID52/gcgdata/geneSeq/geneSeq/NA1994.DAT.*

16: /SID52/gcgdata/geneSeq/geneSeq/NA1995.DAT.*

17: /SID52/gcgdata/geneSeq/geneSeq/NA1996.DAT.*

18: /SID52/gcgdata/geneSeq/geneSeq/NA1997.DAT.*

19: /SID52/gcgdata/geneSeq/geneSeq/NA1998.DAT.*

20: /SID52/gcgdata/geneSeq/geneSeq/NA1999.DAT.*

21: /SID52/gcgdata/geneSeq/geneSeq/NA2000.DAT.*

22: /SID52/gcgdata/geneSeq/geneSeq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result	Query	Match	Length	DB	ID	Description
No.	Score					
1	573	100.0	573	22	AA500039	Human cDNA encodin
2	323	56.4	337	22	AA500040	Human cDNA encodin
3	323	56.4	402	22	AA500041	Human cDNA encodin
4	64.6	11.3	462	22	AA151914	Probe #19/60 used
5	55.8	9.7	114	22	AA151078	Probe #19/60 used
6	44.2	7.7	1525	21	AAAC9899	Human pancreatic c
7	37.8	6.6	2566	21	AAAC6364	Methylolophilus meth
8	37.8	6.6	3984	18	AAAT58627	Human tyrosine p
9	37.4	6.5	717	20	AAAT78463	Human PRPH cDNA.
	37.4	6.5	717	20	AAK24320	Human bone mineral
	4	6.0	717	21	AAAC8713	Human LTM mineral

12	37.4	6.5	1.656	21	AAC8774.1
11	37.4	6.5	1.575	21	AAC8774.2
14	37.4	6.5	1.220	21	AAC8773.7
13	37.4	6.5	1.444	20	AAX2437.2
15	37.4	6.5	1.544	21	AAC8773.5
16	37.4	6.5	1.665	21	AAC8743.8
17	37.4	6.5	1.665	21	AAC8773.8
18	37.4	6.5	1.589	21	AAX2432.2
19	37.4	6.5	1.689	21	AAC8772.2
21	36.6	6.4	4.57	21	AAC7545.5
22	36.6	6.4	4.78	20	AAX2434.3
23	36.2	6.3	1.696	21	AAC8770.4
24	36.2	6.3	1.696	21	AAC8770.4
25	35.8	6.1	3.75	20	AAV8811.7
26	34.8	6.0	7.85	20	AAX2432.2
27	34.2	6.0	7.65	21	AAC8772.2
28	34.2	6.0	1.511	19	AAC8772.2
29	34	5.9	1.593	20	AAZ421.3
30	33.8	5.9	2.400	19	AAV1900.7
31	33.8	5.9	4.768	20	AAZ629.0
32	33.8	5.8	1.597	17	AAV1290.1
33	33.4	5.8	2.162	20	AAX4005.0
34	33.4	5.8	2.236	20	AAX3964.0
35	33.4	5.8	2.236	20	AAV4005.0
36	33.4	5.8	2.289	20	AAV4006.0
37	33.4	5.8	2.309	20	AAV4006.0
38	33.4	5.8	1.336	19	AAV3095.5
39	33.2	5.8	1.554	21	AAV7597.7
40	33.2	5.8	1.864	21	AAV1613.3
41	33.2	5.8	1.844	22	AAH1803.3
42	33.2	5.8	2.519	22	AAH1607.7
43	33.2	5.8	2.581	22	AAH1600.0
44	33.2	5.8	2.987	9	AAAN8045.6
45	33.2	5.8	2.987	9	AAAN8045.6

ALIGNMENTS

RESULT 1

AA500039 standard; cDNA; 573

AA
AC AAS00039

DT 11-MAY-2001 (first entry)

aa Human cDNA encoding novel human protein, NHP#1.
DE

Human: novel human protein; NHP#1; gene therapy; drug screening

obesity; high blood pressure; ss.

yy
OS Homo sapiens

XX	XX	key	Location/Qualifiers
FM			

AA WO200114422-A1
PN

XX
PD 01-MAR-2001

18-AUG-2000; 2000WO-US22815.

XX
PR 24-AUG-1999: 99US-0150511

AA
PA (LEXI-) LEXICON GENETICS INC.

XX
PI
Turner CA, Zambrowicz B, Friedrich G, Nehls M, Sands AT;

XX
25
1997-218430/22

DR P-PSDB; AAU00031

XX	11-MAY-2001	(first entry)	
XX			
XX	XX	Human cDNA encoding novel human protein, NHP#2.	
DE	DE		
XX	XX	Human; novel human protein; NHP#2; gene therapy; drug screening;	
KW	KW	obesity; high blood pressure; ss.	
XX	XX		
XX	XX	Homo sapiens.	
OS	OS		
XX	XX		
XX	Key	Location/Qualifiers	
FF	FF	1.327	
CDS	CDS		
		/*tag= a	
FF	FF	/product= "NHP#2"	
FT	FT		
XX	XX		
XX	XX		
PN	PN	WO200114422-A1.	
XX	XX		
PD	PD	01-MAR-2001.	
XX	XX		
XX	XX		
PF	PF	18-AUG-2000; 2000WO-US22815.	
XX	XX		

XX (LEXI-) LEXICON GENETICS INC.
PA
XX
PI Turner CA, zambrowicz B, Friedrich G, Nehls M, sands AT;
XX

WPI: 2001-218430/22.
P-PSDB: AAU00032.

Novel human polynucleotides isolated from human mammary gland cDNA library, encodes novel human proteins which are useful in diagnosis, drug screening, clinical trial monitoring or treating behavioural disorders.

Claim 3: Page 25; 29pp; English.

The sequence encodes a novel human protein, NHP#2, which shares structural motifs with human APXL protein. Nucleotide constructs encoding functional NHPs are used in gene therapy approaches for the modulation of NHP expression. NHP oligonucleotides can be used as hybridisation probes for screening libraries and assessing NHP gene expression patterns. Also, labeled NHP nucleotide probes can be used to screen a human genomic library. The NHP nucleotide sequences are also useful in drug screening techniques for treating symptomatic or phenotypic manifestations of perturbing the normal function of NHP in the body. Examples of such manifestations may include obesity and high blood pressure.

Sequence 327 BP; 91 A; 76 C; 96 G; 64 T; 0 other;

Query Match	56.4%;	Score 323;	DB 22;	Length 327;
Best Local Similarity	100.0%;	Prod. No. 9.5e-93;		
Matches 323;	Conservative	0;	Mismatches	0;
Indels	0;	Gaps	0;	
1	atgatgaggaccactgaagacttcacaaagcctagtgcacattaaactcaacacgccc	60		
1	atgatgaggaccactgaagacttcacaaagcctagtgcacattaaactcaacacgccc	60		
61	accaagggagggtacattatctgaggcattctctgaggagagagctccctggggtttt	120		
61	accaagggagggtacattatctgaggcattctctgaggagagagctccctggggtttt	120		
121	actctaaagggcgccctggagcaggagaccattaatcatctctaaggtcgagaagggg	180		
121	actctaaagggcgccctggagcaggagaccattaatcatctctaaggtcgagaagggg	180		
181	ggcaaaagacacacccctgagctccaactcgagcctggggatgaggttgtgcacataaat	240		
181	ggcaaaagacacacccctgagctccaactcgagcctggggatgaggttgtgcacataaat	240		
241	gaggtgacttgagcagctccaaagagagcaggtttccctgggtgaaaggagctccacaag	300		

181 ggcaagcagacacccctgagctccaaactgagctggggatgaggttctgcacatcaat 240
 |||||
 181 ggcaagcagacacccctgagctccaaactgagctggggatgaggttctgcacatcaat 240
 |||||
 241 gaggtaactctgaacgactccaaagagagcaggtttcccttggtgaaggatcctacaa 300

Db 241 gaggtgactctgagcagctccagaagagcagcttcctctgtgaaagatcctacaag 300
 QY 301 accctcagcgtgtagtgagcag 323
 Db 301 accctcagcgtgtagtgagcag 323

RESULT 3

AS00041 ID AS00041 standard; cDNA; 402 BP.

AC AAS00041;

DE 11-MAY-2001 (first entry)

XX Human cDNA encoding novel human protein, NHP#3.

XX Human; novel human protein; NHP#3; gene therapy; drug screening;

KW Obesity; high blood pressure; ss.

XX Homo sapiens.

OS Homo sapiens.

FT Key Location/Qualifiers

FT CDS 1..402

FT /tag= "NHP#3"

FT /product= "NHP#3"

XX MO200114422-A1.

XX 01-MAR-2001.

XX 18-AUG-2000; 2000MO-US22815.

XX 24-AUG-1999; 99US-0150511.

XX (LEXI-) LEXICON GENETICS INC.

XX Turner CA, Zambrowicz B, Friedrich G, Nehls M, Sands AT;

XX WPI; 2001-218430/22.

XX P-PSDB; AAU00033.

XX Claim 5; Page 25; 29pp; English.

XX The sequence encodes a novel human protein, NHP#3, which shares

XX structural motifs with human APX1 protein. Nucleotide constructs

XX encoding functional NHPs are used in gene therapy approaches for the

XX modulation of NHP expression. NHP oligonucleotides can be used as

XX hybridisation probes for screening libraries and assessing NHP gene

XX expression patterns. Also, labeled NHP nucleotide probes are also

XX screen a human genomic library. The NHP nucleotide sequences are also

XX useful in drug screening techniques for treating symptomatic or

XX phenotypic manifestations of perturbing the normal function of NHP in the

XX body. Examples of such manifestations may include obesity and high blood

XX pressure.

XX Sequence 402 BP; 110 A; 91 C; 117 G; 84 T; 0 other;

XX Query Match 56.4%; Score 323; DB 22; Length 402;

XX Best Local Similarity 100.0%; Pred. No. 1e-92;

XX Matches 323; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

XX 1 atgatgagacacactgaagactccacaagcctagtcacacattacttaacacgccc 60

XX atgatgagacacactgaagactccacaagcctagtcacacattacttaacacgccc 60

QY 61 accaagggaaggtacattatctctgagggcattcttgaggaggaagctccctgggtttt 120
 Db 61 accaagggaaggtacattatctctgagggcattcttgaggaggaagctccctgggtttt 120
 QY 121 actctaaagggtgctctgagcagcaggaacattatctcttaaggtcgaagaagg 180
 Db 121 actctaaagggtgctctgagcagcaggaacattatctcttaaggtcgaagaagg 180
 QY 181 ggcagaagcagacacccctgagctcccaactcagcagcgtgggagtagttgtcacatcaat 240
 Db 181 ggcagaagcagacacccctgagctcccaactcagcagcgtgggagtagttgtcacatcaat 240
 QY 241 gaggtgactctgagcagctccagaagaggaagcagcttcctctgtgaaagatcctacaag 300
 Db 241 gaggtgactctgagcagctccagaagaggaagcagcttcctctgtgaaagatcctacaag 300
 QY 301 accctcagcgtgtagtgagcag 323
 Db 301 accctcagcgtgtagtgagcag 323

RESULT 4

AI37914 ID AI37914 standard; DNA; 462 BP.

AC AI37914;

DE 17-OCT-2001 (first entry)

XX Probe #6600 used to measure gene expression in human placenta sample.

XX Probe; microarray; human; placenta; antenatal diagnosis;

XX genetic disorder; ss.

XX Homo sapiens.

XX MO200157272-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001MO-US00663.

XX 04-FEB-2000; 2000US-0180312.

XX 26-MAY-2000; 2000US-0207456.

XX 30-JUN-2000; 2000US-0608408.

XX 03-AUG-2000; 2000US-0632366.

XX 21-SEP-2000; 2000US-0234687.

XX 27-SEP-2000; 2000US-0236359.

XX 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-48897/53.

XX Human genome-derived single exon nucleic acid probes useful for

XX analyzing gene expression in human placenta.

XX Claim 25; SEQ ID No 6600; 654pp; English.

XX The present invention relates to single exon nucleic acid probes (SENPs).

XX CC The present sequence is one such probe. The probes are useful for

XX CC producing a microarray for predicting, measuring and displaying gene

XX CC expression in samples derived from human placenta. The probes are useful

XX CC for antenatal diagnosis of human genetic disorders.

XX Query Match 11.3%; Score 64.6; DB 22; Length 462;

XX Best Local Similarity 65.7%; Pred. No. 1.3e-10;

XX Matches 94; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

XX Sequence 462 BP; 109 A; 102 C; 135 G; 116 T; 0 other;

Db	62	gcgaattgccttgtagagggtcccataagaccctgaagctgctgctg
	RESULT	6
	AAC98998	
	ID	AAC98998 standard; cDNA; 1525 BP.
	XX	
	XX	AAC98998;
	XX	
	XX	AC
	XX	(first entry)
	DT	09-MAR-2001
	XX	
	DE	
	XX	Human pancreatic cancer antigen nucleotide sequence SEQ

DT 09-MAR-2001 (first entry)
XX
DE Human pancreatic cancer antigen nucleotide sequence SEQ ID NO:226.

Human; pancreas; pancreatic cancer; pancreatic antigen; detection; diagnosis; identification; cytostatic; neuroprotective; neoplastic; immunomodulatory; relaxant; contraceptive; gynaecological; antiinflammatory; cardiac; gene therapy; chromosome mapping; linkage analysis; tissue identification; tissue typing; forensic; neural; immune system; muscular; reproductive; gastrointestinal; pulmonary; cardiovascular; renal; proliferative; ss.

Homo sapiens.

WO200055320-A1.

21-SEP-2000.

08-MAR-2000; 2000WO-US05989.

12-MAR-1999; 99US-0124270.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Ruben SM;

WPI; 2000-579444/54.

P-PSDB; AAB54233.

New nucleic acid that
treating, or ameliora

cancer, or for use in

Claim 1; Page 669-670

AAC98773 to AAC99231
proteins, called pan-

AAB54466. The human J
neuroprotective, neuro

gynaecological, cardiac, in gene therapy. The

preventing, treating,
for diagnosing a path

subject. Binding partner proteins can be identified

Agonists and antagonists

pancreatic cancer anti-
acid hybridisation pro-

analysis, tissue identification, and diagnostic methods.

which are used to purify both *in vivo* and *in vitro*

proteins can be used
reproductive, gastrof

proliferative disorders
sequences used in the

Sequence 1525 BP; 303

Local Similarity

ches 73; Conservat

CC easier assessment of transfected cells. Clinical applications
 CC include enhancement of bone repair in fractures, bone defects, bone
 CC grafting, and normal homeostasis in patients presenting with
 CC osteoporosis. A method of inducing bone formation using
 CC transfected osteogenic precursor cells is claimed.

XX Sequence 717 BP; 138 A; 262 C; 219 G; 98 T; 0 other;

Query Match

Best Local Similarity 49.8%; Score 37.4; DB 20; Length 717;
 Matches 126; Conservative 0; Mismatches 121; Indels 6; Gaps 1;

QY 86 aggcattcttgaggagagagctccctggtgttacttaaaagggtgctgagacg 145
 DB 14 agtgatgctgagggagagcagcacttggtctccggtgcaagggtgagactta 73
 QY 146 gagaacattatcatctcctaagtcgaaggaggggcaaacagacaccctgagctca 205
 DB 74 atgtgccctctccatttcccgctcactcctggtgggcaaaagcgc-----gcaggccg 127
 QY 206 aactcagcgtgagatgagtggtgacatcaatgagtgactctgagcagctccagaa 265
 DB 128 gaggcgccgtggtgactggtgctgacatgcatgagcagaaatgctggtagctcacac 187
 QY 266 aggaagcaggttccctcgtgaaagagctctacaagaccctcagcgtgtagtgcagcc 325
 DB 188 acatcgaaagctcagaacaagatccggtcgtgagagcgctcagcctggtgctcagca 247
 QY 326 tctccaccagcgt 338
 DB 248 gggcccaagccgt 260

RESULT 11

AAC87713
 ID AAC87713 standard; cDNA; 717 BP.

AC AAC87713;
 XX
 DT 01-MAR-2001 (first entry)

DE Human LIM mineralisation protein (HLMF) partial cDNA SEQ ID NO:7.

XX LIM mineralisation protein; LMP; bone formation; osteopathic;
 KW osteogenic precursor cell; gene therapy; metabolic bone disease;
 KW osteoporosis; bone degenerative disease; ss.

OS Homo sapiens.

PN WO200066178-A1.

PD 09-NOV-2000.

PF 28-APR-2000; 2000MO-US11664.

PR 30-APR-1999; 99US-0132021.

PA (UYEM-) UNIV EMORY.

PI Boden SD, Hair GA;

DR WPI; 2000-672828/65.

XX New nucleic acid encoding a human LIM mineralization protein for
 PT inducing or inhibiting bone formation, fusing a spine, stimulating
 PT production of an osteogenic cell, or treating bone conditions, such as,
 PT osteoporosis

Example 17; Page 61; 84pp; English.

present invention specifically describes the human LIM
 mineralisation proteins (LMP) HLMF-2 and HLMF-3. LMPs have osteopathic

CC activity and can be used in gene therapy. LMP nucleic acids can be used
 CC to induce or inhibit bone formation, fuse a spine, stimulate production
 CC of an osteogenic cell, or inhibit the expression of HLMF-2 or HLMF-3.
 CC They can be used to treat bone conditions, such as, osteoporosis and
 CC other metabolic bone diseases. Antibodies to the LMP proteins encoded
 CC by the nucleic acids are used in marker assays to identify risk factors
 CC in bone degenerative diseases, such as osteoporosis. The nucleic acids
 CC are used in gene therapy for bone formation which leads to the
 CC advantages of: (1) lower production costs; (2) greater efficacy compared
 CC to extracellular treatment regimens due to the ability to achieve
 CC prolonged expression of the intracellular signal; (3) by-passing the
 CC possibility that treatment with extracellular signals might be hampered
 CC due to the presence of limiting numbers or receptors for those signals;
 CC (4) permitting the delivery of transfected potential osteoprogenitor
 CC cells directly to the site where localised bone formation is required;
 CC and (5) permitting systemic bone formation, which provides a treatment
 CC regimen for osteoporosis and other metabolic bone diseases. The present
 CC sequence represents a partial human LMP cDNA sequence, which is given
 CC in an example from the present invention.

XX Sequence 717 BP; 138 A; 262 C; 219 G; 98 T; 0 other;

Query Match

Best Local Similarity 49.8%; Score 37.4; DB 21; Length 717;
 Matches 126; Conservative 0; Mismatches 121; Indels 6; Gaps 1;

QY 86 aggcattcttgaggagagagctccctggtgttacttaaaagggtgctgagacg 145
 DB 14 agtgatgctgagggagagcagcacttggtctccggtgcaagggtgagactta 73
 QY 146 gagaacattatcatctcctaagtcgaaggaggggcaaacagacaccctgagctca 205
 DB 74 atgtgccctctccatttcccgctcactcctggtgggcaaaagcgc-----gcaggccg 127
 QY 206 aactcagcgtgagatgagtggtgacatcaatgagtgactctgagcagctccagaa 265
 DB 128 gaggcgccgtggtgactggtgctgacatgcatgagcagaaatgctggttagctcacac 187
 QY 266 aggaagcaggttccctcgtgaaagagctctacaagaccctcagcgtgtagtgcagcc 325
 DB 188 acatcgaaagctcagaacaagatccggtcgtgagagcgctcagcctggtgctcagca 247
 QY 326 tctccaccagcgt 338
 DB 248 gggcccaagccgt 260

RESULT 12

AAC87741
 ID AAC87741 standard; cDNA; 1456 BP.

AC AAC87741;

DT 01-MAR-2001 (first entry)

DE Human LMP-2 (HLMF-2) encoding cDNA sequence SEQ ID NO:37.

XX LIM mineralisation protein; LMP; bone formation; osteopathic;
 KW osteogenic precursor cell; gene therapy; metabolic bone disease;
 KW osteoporosis; bone degenerative disease; ss.

OS Homo sapiens.

PN WO200066178-A1.

PD 09-NOV-2000.

PF 28-APR-2000; 2000MO-US11664.

PR 30-APR-1999; 99US-0132021.

PA (UYEM-) UNIV EMORY.

